

**Görüntü Ayırıştırma için Özgün Süperpiksel Bölütleme  
Algoritmaları Geliştirilmesi**

**Program Kodu: 1001**

**Proje No: 115E307**

Proje Yürütücüsü:  
**Prof. Dr. Hasan F. ATEŞ**

Bursiyerler:

Sercan SÜNETCİ

Elçin DEĞİRMENCİ

EYLÜL 2017  
İSTANBUL

## Önsöz

TÜBİTAK Bilimsel ve Teknolojik Araştırma Projelerini Destekleme Programı (no. 1001) bünyesinde desteklenmiş ve Işık Üniversitesi'nde yürütülmüş olan bu projede, bilgisayarla görünün önemli problemlerinden görüntü ayrıştırma / sahne etiketleme için özgün yöntemler geliştirilmiştir. Yapılan akademik çalışmalar, ülkemizde bilgisayarla görü alanında sahip olunan bilgi birikimini artırmış, makale ve konferanslar aracılığıyla akademik çevreler tarafından farkedilmiş, ve çeşitli ulusal projelere katkı sağlamıştır.

Bu projenin başarıyla yürütülmesinde, TÜBİTAK'a sağladığı destek için teşekkür ederiz.

# İçindekiler

Önsöz.....	ii
İçindekiler .....	iii
Tablo Listesi .....	iv
Şekil Listesi .....	v
Özet .....	vi
Abstract.....	vii
1. GİRİŞ .....	1
1.1 Literatür Özeti .....	2
1.2 Projenin İçeriği .....	5
2. SÜPERPİKSEL TABANLI GÖRÜNTÜ AYRIŞTIRMA .....	7
2.1 SuperParsing Algoritması.....	7
2.2 SuperParsing Performans Analizi.....	10
3. SUPERPARSING ALGORİTMASININ İYİLEŞTİRİLMESİ.....	12
3.1 Çok-hipotezli Süperpiksel Bölütleme ile Görüntü Ayrıştırma.....	12
3.2 SuperParsing’de Kernel-tabanlı Olabilirlik Kestirimi.....	14
3.2.1 Benzetimler.....	15
3.2.2 Çok-hipotezli SuperParsing’de Ağırlıklı Olabilirlik Kestirimi .....	18
3.3 Süperpiksel Bölütleme ve Görüntü Ayrıştırma Ortak Eniyileme .....	19
3.4 Süperpiksel Öznitelikleri için Kodlama Yöntemleri .....	24
3.4.1 Öznitelik Kodlama Yöntemleri .....	25
3.4.2 Benzetimler.....	28
3.5 CNN Özniteliklerinin SuperParsing’de Kullanılması.....	31
4. ÇOK-HİPOTEZLİ BAĞLAMSAL MODELLEME .....	35
4.1 Alternatif Bölütlemeler için MRA Bağlamsal Modeli .....	35
4.2 Çok-hipotezli MRA için Benzetimler .....	38
4.2.1 Derin Ağlarda Bağlamsal Çıkarsama .....	43
5. UZAKTAN ALGILAMA İÇİN GÖRÜNTÜ AYRIŞTIRMA .....	47
5.1 ISPRS 2-Boyutlu Anlambilimsel Etiketleme Yarışması .....	48
5.2 Benzetimler.....	50
6. SONUÇ VE ÖNERİLER .....	53
REFERANSLAR.....	55

## Tablo Listesi

Tablo 2.1. Farklı bölütleme yöntemleri için SuperParsing ayrıştırma doğrulukları .....	11
Tablo 3.1. SIFT Flow veri kümesi için piksel etiketleme doğruluğu .....	16
Tablo 3.2. Farklı sınıflar için sınıflandırma doğrulukları.....	17
Tablo 3.3. Farklı öznitelikler için sınıflandırma doğrulukları.....	20
Tablo 3.4. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ..	29
Tablo 3.5. SIFT betimleyici için doğruluk (piksel/sınıf ort.) karşılaştırması .....	30
Tablo 3.6. SIFT Flow veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları ( $K=400, S=200, R=15$ ).....	33
Tablo 3.7. SIFT Flow veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları ( $K=200, S=100, R=30$ ).....	33
Tablo 3.8. LabelMe veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları ( $K=200, S=100, R=30$ ).....	34
Tablo 4.1. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ..	40
Tablo 4.2. LabelMe veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ....	42
Tablo 4.3. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ..	44
Tablo 4.4. LabelMe veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ....	45
Tablo 5.1. Vaihingen veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması ..	52

## Şekil Listesi

Şekil 1.1. SLIC süper-pikselleri (Achanta vd., 2012) (boyut: 64, 256, 1024 piksel (yaklaşık))	2
Şekil 1.2. Süperpiksel bölütleme.....	3
Şekil 3.1. Görsel karşılaştırma (mavi:bina, açık mavi:kapı, kahverengi:pencere, turuncu: kaldırım, sarı: yol).....	13
Şekil 3.2. Görsel karşılaştırma (kırmızı: gökyüzü, yeşil: dağ, açık yeşil: arazi, açık turuncu: yol, turuncu: kum).....	18
Şekil 3.3. Farklı öznitelikler için etiketleme doğruluğu – karşılıklı bilgi miktarları karşılaştırması (( $K=300$ , $S=150$ ) bölütleme için). ....	22
Şekil 3.4. Görsel karşılaştırma (kırmızı: gökyüzü, yeşil: dağ, k.rengi: ağaç, mavi: yeşil alan, sarı: bitki) .....	30
Şekil 4.1. Çok-hipotezli MRA modeli.....	36
Şekil 4.2. Görsel karşılaştırma: (a) Orijinal görüntü; (b) Doğru etiketleme; (c) SuperParsing; (d) DC3. ....	42
Şekil 4.3. FCN bölütleme: (a) Orijinal görüntü; (b) FCN etiketleri; (c) FCN süperpikselleri. ...	43
Şekil 4.4. Görsel karşılaştırma: (a) Orijinal görüntü; (b) Doğru etiketleme; (c) $SP1$ ; (d) $SP2$ ; (e) DC2.....	46
Şekil 5.1. Vaihingen ortofoto mozaïği.....	49
Şekil 5.2. Vaihingen eğitim verisi: (a) 3-bant multispektral görüntü; (b) DSM görüntüsü; (c) Doğru etiketleme. ....	49
Şekil 5.3. Vaihingen yükseklik verisi: (a) 3-bant multispektral görüntü; (b) DSM görüntüsü; (c) NDSM (Düzgelenmiş yükseklik) görüntüsü.....	50
Şekil 5.4. Görsel sonuç (beyaz: yüzey, mavi:bitki örtüsü, lacivert: bina, yeşil: ağaç, sarı: araba): (a) 3-bant multispektral görüntü; (b) Doğru etiketleme; (c) DC2 ( $SP2$ , $SP4$ ) sonucu. ....	52

## Özet

Süperpikseller görüntü bölütleme ve ayrıştırma problemlerinde yaygın olarak kullanılmaktadır. Sahne etiketlemede görüntü bir süperpiksel algoritması ile görsel olarak tutarlı küçük parçalara bölütlenmekte; daha sonra süperpikseller farklı sınıflara ayrıştırılmaktadır. Bu projede bölütleme ve etiketleme bütünsel bir bakış açısı ile ele alınarak görüntü ayrıştırmanın farklı adımları için özgün yaklaşımlar geliştirilmiştir. Yapılan çalışmalar, süperpikseller için alternatif bölütleme, öznitelik çıkarımı, sınıf-olabilirlik hesaplama ve bağlamsal modelleme yöntemleri geliştirilmesini kapsamaktadır. Projede öncelikle farklı bölütleme yöntem ve parametrelerinin etiketleme doğruluğu üzerindeki etkisi test edilmiştir. Daha sonra süperpiksel özniteliklerinin seçimi ve kodlanması, sınıf etiketlerinin olabilirlik hesabının modellenmesi üzerinde durulmuştur. Son olarak, alternatif bölütleme sonuçlarının kaynaştırılması için genelleştirilmiş bağlamsal modelleme yaklaşımı geliştirilmiştir. Önerilen yöntemler çeşitli anlambilimsel görüntü veritabanlarında test edilmiş ve eniyilenmiştir. Ayrıca projenin son döneminde, yapılan çalışmalar uydu görüntülerinden arazi örtüsü sınıflandırma problemine uyarlanmıştır. Benzetim sonuçları, farklı bölütleme yöntemlerindeki tümler bilginin doğru şekilde birleştirilmesiyle görüntü etiketleme doğruluğunda ciddi artışlar elde edildiğini ortaya koymuştur.

**Anahtar Kelimeler:** görüntü ayrıştırma, etiketleme, bölütleme, süperpiksel.

## Abstract

Superpixels are widely used in image segmentation and parsing problems. In scene labeling, image is initially divided into visually consistent small pieces by using a superpixel algorithm; later superpixels are parsed into different classes. In this project, segmentation and labeling are considered together in a global perspective, and novel approaches are proposed for different steps of image parsing. In particular, several methods are developed for alternative segmentation, feature extraction, class-likelihood computation and contextual modeling of superpixels. Initially the effect of different segmentation methods and parameters on the labeling accuracy is thoroughly tested. Later superpixel feature selection and coding, modeling of likelihood computation for class label likelihoods are investigated. Finally a generalized contextual modeling framework is developed for the fusion of alternative segmentation results. The proposed methods are tested and optimized on several semantic image databases. In addition, in the final phase of the project, this work is adapted for the problem of land cover classification from satellite images. Simulation results show that it is possible to achieve substantial improvement in image labeling accuracy by accurate combination of complementary information coming from different segmentation methods.

**Keywords:** image parsing, labeling, segmentation, superpixel.

# 1. GİRİŞ

Görüntü ve nesne sınıflandırma üzerinde uzun yıllardır çalışılan bilgisayarla görünün en temel problemlerindendir (Torralba vd., 2008; Lowe, 1999). Son on yılda makina öğrenme alanında yaşanan gelişmeler ve paralel programlama ile artan işlem kapasitesi sayesinde, özellikle karmaşık güdümlü sınıflandırma yöntemlerinin uygulanabilir hale geldiği ve sınıflandırma doğruluğunda ciddi artışlar elde edildiği gözlenmektedir (van de Sande vd., 2010).

Görüntü ayrıştırma (ya da, anlambilimsel sahne bölütleme) bir sahnenin farklı bileşenlerine bölütlenmesi ve bu bileşenlerin etiketlenmesi olarak tanımlanabilir (Farabet vd., 2013). Görüntü ayrıştırma, sahnede varolan nesnelere ve yerlerini tespit etmeyi sağladığı için sahnenin içeriğinin anlamlandırılmasına imkan sağlamakta ve pek çok yapay zeka uygulaması için önem taşımaktadır.

Süperpikseller görüntü bölütleme ve ayrıştırma problemlerinde yaygın olarak kullanılmaktadır (Nguyen vd., 2016; George, 2015; Tighe ve Lazebnik, 2013a). Süperpiksel tabanlı bölütlemede görüntü görsel olarak anlamlı atomik bölgelere ayrılarak nesne kenarlarıyla uyumlu çok sayıda bölüt elde edilmektedir. Süperpikseller, nesne sınırlarıyla tutarlı olmanın yanında, görüntüdeki artıklığı temsil edebilmeleri ve özgün bölüt-tabanlı öznitelikler tanımlanmasına imkan vermeleri sayesinde pek çok bilgisayarlı görü problemi için cazip bir gösterim sunmaktadır (bkz. Şekil 1.1). Sahne etiketlemede görüntü öncelikle bir süperpiksel algoritması ile görsel olarak tutarlı küçük parçalara bölütlenmekte; daha sonra süperpikseller farklı sınıflara ayrıştırılmaktadır.

Bu projede özgün süperpiksel tabanlı bölütleme ve ayrıştırma yöntemleri geliştirilmesi üzerinde durulmuştur. Literatürden farklı olarak, bölütleme ve sınıflandırma problemlerini ortak ele alan bir yaklaşım ile hem bölütleme hem de sınıflandırma başarımının birlikte eniylenmesi hedeflenmiştir. Proje süresince görüntü ayrıştırma ile ilgili aşağıda sıralanan başlıklarda çalışmalar yürütülmüştür:

1. Farklı bölütleme yöntem ve parametrelerinin ayrıştırma başarımı üzerindeki etkisi incelenmiştir.
2. Süperpiksel bölütleri için çeşitli öznitelikler tanımlanmış ve test edilmiştir. Bölütler için şekil, geometri, renk ve doku tabanlı öznitelikler yanında “derin sinir ağlarında” öğrenilmiş öznitelikler kullanılmıştır. Ayrıca öznitelik vektörleri için çeşitli kodlama yöntemleri denenmiştir.



3. Bölütlerin etiketlenmesinde farklı sınıflar için koşullu olasılık dağılımlarının modellenmesi ve olabilirlik skorlarının tespitinde özgün yaklaşımlar geliştirilmiştir.
4. Alternatif bölütleme ve ayrıştırma yöntemlerinin sonuçlarını kaynaştırmak için genelleştirilmiş bağlamsal çıkarsama modeli önerilmiştir. Bu model ile birden fazla bölütleme hipotezindeki tümleyici bilginin eniyi şekilde birleştirilmesinin görüntü etiketleme doğruluğunu iyileştirdiği gösterilmiştir. Bölütleme için süperpiksel tabanlı yöntemlerle birlikte “derin” evrimsel sinir ağlarının da kullanılmasıyla literatürdeki en son etiketleme doğruluklarının üzerine çıkılmıştır.
5. Önerilen yöntemler çeşitli anlambilimsel görüntü veri-tabanlarında test edilmiş ve eniyelenmiştir. Ayrıca bu yöntemler uydu görüntülerinden arazi örtüsü sınıflandırma problemine uyarlanmıştır.



Şekil 1.1. SLIC süper-pikselleri (Achanta vd., 2012) (boyut: 64, 256, 1024 piksel (yaklaşık))

Bir sonraki bölümde, bu proje çıktılarının görüntü ayrıştırma literatüründeki yeri, son yıllarda bu alanda yapılan çalışmalarla karşılaştırılarak detaylı şekilde açıklanmıştır.

## 1.1 Literatür Özeti

Görüntü ayrıştırma, sahnedeki her bir pikselin hangi görsel/anlambilimsel/geometrik sınıfa ait olduğunun belirlenmesi problemidir. Görüntü ayrıştırma, bölütleme, nesne sezimi ve çok-sınıflı etiketleme problemlerini içinde barındırdığı için oldukça zor ve gelişmeye açık bir araştırma konusudur (Farabet vd., 2013). Literatürde birçok farklı görüntü ayrıştırma / etiketleme metodu önerilmiştir; bunlardan bazıları piksel piksel ayrıştırma yaparken (Shotton vd., 2008), bazıları ise görüntü bölütleri üzerinde öznitelikleri birleştirerek ayrıştırma yapmaktadır (Gould vd., 2009).

Görüntünün bölütlenmesinde son yıllarda pek çok süperpiksel tabanlı yöntem önerilmiştir. Piksel tabanlı gösterime göre daha esnek ve daha tutarlı olması sayesinde,

süperpiksel tabanlı bölütleme, sınıflandırma ve nesne sezim algoritmalarında yüksek başarımlı göstermiştir (Fulkerson vd., 2009; Kluckner vd., 2010). Süperpikseller, görüntüyü aşırı bölütleyerek (oversegmentation) kendi içinde tutarlı çok sayıda bölgeye ayırmaktadır (Achanta vd., 2012; Levinshtein vd., 2009) (bkz. Şekil 1.1 ve Şekil 1.2). Daha sonra öznelik vektörleri, piksel seviyesinde hesaplanmak yerine, her bir süper-piksel için çıkarılmakta ve sınıflandırma piksel seviyesi yerine süperpiksel seviyesinde yürütülmektedir. Süperpikseller, nesne sınırlarıyla tutarlı olmanın yanında, görüntüdeki artıklığı temsil edebilmeleri ve özgün bölüt-tabanlı öznelikler tanımlanması imkan vermeleri sayesinde pek çok bilgisayarlı görü problemi için cazip bir gösterim sunmaktadır (Achanta vd., 2012).



Şekil 1.2. Süperpiksel bölütleme

Literatürde pek çok süperpiksel bölütleme yöntemi mevcuttur (Achanta vd., 2010; Levinshtein vd., 2009; Ming-Yu vd., 2011; Vedaldi ve Soatto, 2008). Bu yöntemler, ortaya çıkan süperpiksellerin boyutu, düzenliliği gibi model kısıtlarıyla birbirinden ayrılmaktadır (Ming-Yu vd., 2011). Süperpiksellerin nesne sınıflandırma ve görüntü bölütlemeye kullanılmasında SIFT gibi standart özneliklerin yanında her bir bölütün şekilsel/morfolojik ve geometrik özelliklerine dayanan farklı özneliklerden de faydalandığı anlaşılmaktadır (Kluckner vd., 2010; Aksoy, 2006).

Süperpikselleri görüntü ayrıştırma için kullanan çalışmalar giderek artmıştır (Razzaghi ve Samavi, 2014). Görüntü ayrıştırmada SuperParsing algoritması (Tighe ve Lazebnik, 2010, 2013a) bir milat olmuş, ve daha sonra pek çok çalışma bu algoritmayı örnek almış ve etiketleme başarımını geçmeye çalışmıştır (Nguyen vd., 2015; Farabet vd., 2013; Eigen ve Fergus, 2012). SuperParsing, farklı bölüt-tabanlı özneliklerin tanımlanması ve başarımlarının karşılaştırılması konusunda ilk çalışmalardan birisidir. Sonraki çalışmalar

daha karmaşık olasılıksal modelleme ve sınıflandırıcılar kullanarak etiketleme doğruluğunu arttırmıştır (George, 2015; Tighe vd., 2015; Eigen ve Fergus, 2012).

SuperParsing ve benzeri yöntemler parametrik olmayan ve sınıflandırıcı eğitimi gerektirmeyen veri-güdümlü bir yaklaşımı benimsemiştir. Bu yaklaşımda test görüntüsünün süperpikselleri eğitim kümesindeki süperpiksellerle eşleştirilerek süperpiksel sınıf etiketleri kestirilmeye çalışılmaktadır (Tighe ve Lazebnik, 2013a; Razzaghi ve Samavi, 2014). Diğer bazı çalışmalarda ise gelişmiş öğrenme teknikleri kullanılarak süperpiksellerin etiketlendiği görülmektedir (Nguyen vd., 2016; Farabet vd., 2013).

Son yıllarda derin Evrişimsel Sinir Ağlarının (Convolutional Neural Network (CNN)) bilgisayarla görü problemlerindeki başarısıyla paralel olarak, anlambilimsel sahne bölütleme için CNN mimarisini kullanan çalışmaların yaygınlaştığı görülmektedir (Liu vd., 2016; Liang vd., 2015; Sharma vd., 2014; Shuai, 2016a, 2016b). Büyük görüntü veri tabanlarında eğitilmiş derin mimarilerden faydalanan bu çalışmalar ile sahne etiketleme başarımında kayda değer artışlar sağlanmıştır. Tam Evrişimsel Ağ (Fully Convolutional Network (FCN)) (Shelhamer vd., 2016) mimarisi ile piksel seviyesinde etiketleme yüksek doğruluk oranlarıyla gerçekleştirilebilmiştir. Bu çalışmaların ortaya çıkışıyla süperpiksel-tabanlı yöntemlerin popülerliği azalmış olmakla birlikte süperpiksellerin sınıflandırılmasında CNN mimarisini kullanan çalışmalar da bulunmaktadır (Nguyen vd., 2016; Mostajabi vd., 2015).

Süperpikseller renk açısından benzer komşu pikselleri gruplayarak belli bir uzamsal tutarlılığı garanti etmekle birlikte, görüntü sınıflandırma ve ayrıştırma problemlerinde süperpiksel komşuluklarının modellenmesinin sınıflandırma başarımını artırdığı görülmektedir. Farklı sınıflar arasındaki komşuluk ilişkilerini yansıtan bağlamsal modelleme için literatürde genelde Markov Rastgele Alanlar (MRA) ya da Koşullu Rastgele Alanlar (KRA) tabanlı modeller kullanılmıştır (Tighe ve Lazebnik, 2013b; Farabet vd., 2013). Kullanılan çeşitli enerji modelleri ile, süperpiksellerin birbirinden bağımsız etiketlenmesiyle karşılaştırıldığında sınıflandırma doğruluğunda kayda değer artışlar elde edilmiştir (Yang vd., 2014; Tighe ve Lazebnik, 2013b). KRA modellerinin derin ağlarda da kullanıldığı çalışmalar literatürde mevcuttur (Liang-Chieh vd., 2017; Zheng vd., 2015).

Bu projede öncelikle SuperParsing algoritmasının ayrıştırma başarımının iyileştirilmesi üzerinde durulmuştur. Bunun için literatürde varolan çalışmalar, süperpikseller arası eşleştirmeyi iyileştirme (Nguyen vd., 2015), uyarlanabilir olabilirlik hesabı (Eigen ve Fergus, 2012), gelişmiş betimleyiciler ve sınıflandırıcılar kullanma (George, 2015; Yang vd., 2014) olarak özetlenebilir. Biz de projede farklı olabilirlik modelleri, süperpiksel öznitelikleri ve görüntü içeriğine göre uyarlanabilir yaklaşımlar geliştirdikten sonra, literatürden farklı olarak, birden fazla bölütleme ya da ayrıştırma sonucunun eniyi şekilde kaynaştırılması için özgün bir yöntem üzerinde durduk. Bunun için önerdiğimiz MRA-tabanlı bağlamsal modellemenin

sadece SuperParsing’de değil herhangi farklı bölütleme/ayırıştırma yöntemlerinin sonuçlarının birleştirilmesi için kullanılabileceğini gösterdik. Benzetimler sonucunda SuperParsing ve FCN bölütleme kararlarının kaynaştırılması ile literatürdeki sonuçlardan daha iyi ayırıştırma doğruluğu elde ettik.

Sonuç olarak, literatüre bakıldığında, süperpiksel bölütleme ve sahne etiketlemenin birbirinden bağımsız iki problem olarak ele alındığı görülmektedir. Ne var ki türetilen süperpiksellerin boyut, şekil ve doku özelliklerinin etiketleme doğruluğu üzerinde önemli etkisi vardır. Süperpiksellerin boyut, renk dağılımı, doku gibi çeşitli özelliklerinin farklı sınıflardaki etiketleme doğruluğuna etkisini inceleyen bir çalışmaya literatürde rastlanmamıştır. Bu projenin katkılarından birisi literatürdeki bu boşluğu doldurmak olmuştur.

Proje çıktılarının literatüre en önemli katkısı ise alternatif süperpiksel bölütleme yöntem ve parametrelerinin ayırıştırma başarımı üzerindeki etkilerini dikkate alarak bu alternatiflerin eniyi şekilde kaynaştırılmasını sağlayan bir bağlamsal model geliştirilmesi olmuştur. Dolayısıyla sınıflandırıcı, bir ön-işlem ile belirlenmiş süperpikselleri etiketlemek yerine, farklı bölütleme ve etiketleme sonuçlarını birlikte değerlendirerek ortak bir sonuca varmaktadır. Bölütleme ve sınıflandırma problemlerini ortak ele alan bu yaklaşım sayesinde hem bölütleme hem de sınıflandırma başarımının birlikte eniyilenmesi hedeflenmiştir.

Projenin son döneminde geliştirilen yöntemler uzaktan algılama uygulamalarında test edilmiştir. Süperpiksellerin uzaktan algılama problemlerinde kullanılması henüz oldukça yeni bir uygulama alanıdır. Arazi örtüsü sınıflandırma (Kluckner vd., 2010) ve değişiklik sezimi problemlerinde (Wu vd., 2012; Gadhiraaju vd., 2014) süperpiksel kullanan bazı çalışmalar bulunmaktadır. Fakat bu çalışmalar genelde standart süperpiksel bölütleme ve güdümlü/güdümsüz sınıflandırma yöntemlerinin ilgili uzaktan algılama problemine uyarlanmasının ötesine geçmemektedir. Stefanski vd. (2013) arazi örtüsü etiketleme için süperpiksel bölütleme parametrelerinin çapraz-sağlama ile eniyilenmesini önermişlerdir. Fakat bu makalede tek bir süperpiksel yöntemi üzerinde durulmaktadır ve spektral değerler dışında herhangi bir betimleyici denenmemiştir. Bu projede ise bölütleme kriterlerinin, süperpiksel betimleyicilerin ve bağlamsal modellemenin arazi örtüsü sınıflandırma başarımını eniyileyecek şekilde seçimi üzerinde durulmuştur.

## **1.2 Projenin İçeriği**

Bölüm 1.1’de genel hatlarıyla özetlenen proje çalışmaları bu sonuç raporunun ilgili bölümlerinde detaylı olarak ele alınmıştır. Bölüm 2’de bu proje çalışmaları için temel teşkil eden SuperParsing algoritması (Tighe ve Lazebnik, 2013a) detaylı bir şekilde açıklanmaktadır. Bölüm 2.2’de SuperParsing farklı süperpiksel bölütleme yöntemleri ve

parametre ayarları ile test edilmiş ve bölütlemenin görüntü ayrıştırma başarımı üzerindeki etkisi analiz edilmiştir.

Bölüm 3'te SuperParsing'in farklı adımları için önerdiğimiz iyileştirmeler ele alınmıştır. Bölüm 3.1 proje ön çalışması olan çok-hipotezli süperpiksel bölütleme ve ayrıştırımadan bahsetmektedir. Bölüm 3.2'de süperpikselerin etiketlenmesinde sınıf-koşullu olabilirlik oranlarının kestirimi için özgün bir yaklaşım sunulmuştur. Bölüm 3.3'te bölütleme ve ayrıştırma ortak eniyileme için çeşitli uyarlanabilir yöntemler önerilmiştir. Bölüm 3.4 süperpikseller için öznitelik seçimi ve kodlanması üzerinde durmaktadır. Bölüm 3.5 derin evrimsel sinir ağlarında öğrenilmiş özniteliklerin süperpikseller için kullanılmasını ele almaktadır.

Bölüm 4'te bu projenin birincil amacı olan bölütleme ve ayrıştırma ortak eniyileme için önerdiğimiz çok-hipotezli bağlamsal modelleme yöntemi açıklanmaktadır. Süperpiksel komşuluklarının birden fazla bölütleme üzerinden tanımlandığı bu modelde farklı ayrıştırma yöntemleri ve kararları kaynaştırılarak daha yüksek etiketleme doğruluğu elde edilebildiği gösterilmiştir. Bölüm 4.1'te bu genelleştirilmiş MRA bağlamsal modelinin teorik altyapısı sunulduktan sonra Bölüm 4.2'de gerçekleştirilen algoritma, modeller ve benzetimlerle ilgili detaylara yer verilmiştir. Bölüm 4.2.1'de ise evrimsel sinir ağları tabanlı FCN yönteminin bu model içerisinde nasıl kullanılabileceği açıklanmış ve ilgili benzetimler sunulmuştur.

Bölüm 5'te, geliştirilen yöntemlerin uzaktan algılama için uydu görüntülerinden arazi örtüsü sınıflandırma problemine uyarlanmasından bahsedilmektedir. Standart süperpiksel öznitelikleri yanında yükseklik verisinin de kullanılmasının sınıflandırma başarımına etkisi Bölüm 5.2'de analiz edilmiştir.

Rapor, proje sonuçlarının tartışıldığı ve ileride yapılması düşünülen çalışmaların ele alındığı sonuç bölümüyle son bulmaktadır.

## 2. SÜPERPİKSEL TABANLI GÖRÜNTÜ AYRIŞTIRMA

Bu bölümde iyileştirilmesi hedeflenen süperpiksel-tabanlı SuperParsing yaklaşımı, algoritma bileşenleri (bölütleme, öznitelik çıkarımı, eğitim alt-seti belirleme, vs.) ve bölütleme parametrelerinin ayrıştırma başarımı üzerindeki etkileri, çeşitli benzetimlerle detaylı şekilde açıklanmıştır.

Sahne etiketlemede görüntü öncelikle bir süperpiksel algoritması ile görsel olarak tutarlı küçük parçalara bölütlenmekte; daha sonra süperpikseller farklı sınıflara ayrıştırılmaktadır. Süperpikseller, görüntüyü aşırı bölütleyerek kendi içinde tutarlı çok sayıda bölgeye ayırmaktadır. Daha sonra öznitelik vektörleri, piksel seviyesinde hesaplanmak yerine, her bir süperpiksel için çıkarılmakta ve sınıflandırma piksel seviyesi yerine süperpiksel seviyesinde yürütülmektedir.

### 2.1 SuperParsing Algoritması

Görüntü ayrıştırma için temel alınan yaklaşım SuperParsing algoritmasıdır (Tighe ve Lazebnik, 2010, 2013a). Süperpiksel tabanlı etiketlemenin yüksek başarımını ortaya koyan ilk çalışma SuperParsing algoritması olmuştur. SuperParsing, görüntü ayrıştırma için parametrik olmayan, veri güdümlü bir yaklaşımı kullanmaktadır. Bu tarz yaklaşımlarda sınıflandırıcı eğitimi yoktur; bunun yerine her yeni test görüntüsü için en benzer eğitim görüntüleri belirlenir ve sınıflandırma için gerekli bilgi eğitim görüntülerinden türetilir. SuperParsing algoritması görüntü etiketleme için süperpikselleri kullanan, eğitim gerektirmeyen ve kolaylıkla binlerce görüntü ve yüzlerce etiket içeren büyük veri kümeleri için de ölçeklenebilen bir yöntemdir.

SuperParsing'de test görüntüsü hakkında bir değerlendirme yapabilmek için öncelikle benzer sahneye sahip görüntülerden bir "geri erişim kümesi" belirlenir. İyi bir geri erişim kümesi, test görüntüsüne benzer nesnelere ve uzamsal dağılım içeren sahne tiplerine sahip görüntülerden oluşmalıdır. Geri erişim kümesinin tespitinde global görüntü betimleyicileri kullanılarak sahne seviyesinde eşleştirme yapılır; daha sonra yerel öznitelikler kullanılarak süperpiksel seviyesinde eşleştirme ile görüntü etiketlenir.

#### Geri Erişim Kümesi

SuperParsing'de bir test görüntüsünü ayrıştırmak için ilk adım eğitim görüntülerinden oluşan bir geri erişim alt-kümesi bulmaktır. Bu görüntüler süperpiksel seviyesinde etiketleme yapabilmek için kullanılacaktır. İyi bir geri erişim kümesi, test görüntüsüne benzer nesnelere

ve uzamsal dağılım içeren sahne tiplerine sahip görüntülerden oluşmalıdır. Bu tarz bir benzerliği yakalayabilmek için, 3 adet global öznitelik kullanılmıştır: *uzamsal piramit*, *gist* (Oliva ve Torralba, 2006) ve *renk histogramı*. Her bir öznitelik tipi için, tüm eğitim görüntülerinin test görüntüsüne olan Euclid ( $L_2$ ) uzaklığı hesaplanarak küçükten büyüğe sıralanır. Daha sonra görüntüler üç listedeki en düşük sıra numaralarına göre tekrar sıralanır ve ilk 200 görüntü belirlenerek geri erişim kümesi olarak atanır.

### Süperpiksel Öznitelikleri

Sorgulama görüntülerinin geri erişim kümesinin içeriğine bağlı bir şekilde etiketlenmesinde süperpikseller kullanılmaktadır. Böylelikle hem hesaplama karmaşıklığı azalmakta hem de öznitelikler uzamsal olarak daha tutarlı bölgeler için hesaplanmaktadır. SuperParsing'de süperpiksel bölütleme için çizge tabanlı bölütleme (ÇTB) algoritması (Felzenszwalb ve Huttenlocher, 2004) kullanılmıştır. Etiketleme için süperpiksel seviyesinde 20 farklı öznitelik kullanılmaktadır. Bu öznitelikler süperpikselle ilgili şekil (ör. süperpiksel alanı), konum (ör. süperpiksel maskesi), doku (ör. tekston / SIFT histogramları), renk (ör. renk histogramı) ve görünüm (ör. gri-seviye gist (Oliva ve Torralba, 2006)) özelliklerini tanımlamaktadır. Sınıflandırma açısından önemli olan tekston ve yoğun SIFT histogram betimleyicileri süperpiksel bölgesi üzerinde ve bu bölgenin 10 piksel genişletilmiş hali üzerinde hesaplanmıştır. Ayrıca SIFT öznitelikleri için süperpikselin sol, sağ, üst, alt sınırlarında farklı histogramlar hesaplanmıştır. Eğitim kümesindeki tüm süperpikseller için bu öznitelikler hesaplanır ve daha sonra bunlar sınıf etiketleriyle birlikte depolanır. Eğer süperpikselin %50 veya daha fazla kısmı bir etiketin bölüt maskesi ile örtüşüyorsa, sınıf etiketi o eğitim süperpikseline atanmaktadır.

### Yerel Süperpiksellerin Etiketlenmesi

Test görüntüleri bölütlenip, geri erişim kümeleri tespit edilip, tüm süperpikseller için öznitelikleri hesaplandıktan sonra sıra her bir süperpikselin tüm olası sınıflar için olabilirlik oranlarının belirlenmesindedir. SuperParsing'de her süperpiksel için log-olabilirlik oranları, geri erişim kümesindeki süperpiksel öznitelikleri ile eşleştirme yapılarak hesaplanmaktadır. Özniteliklerin birbirinden bağımsız olduğu varsayımı altında her bir sınıf  $c$  ve süperpiksel  $s_i$  için logaritmik olabilirlik oranı şöyledir:

$$L(s_i, c) = \log \frac{P(s_i|c)}{P(s_i|\bar{c})} = \sum_k \log \frac{P(f_i^k|c)}{P(f_i^k|\bar{c})} \quad (2.1)$$

Burada  $\bar{c}$ ,  $c$  dışındaki tüm sınıfların kümesi, ve  $f_i^k$ ,  $s_i$  için  $k$ . öznitelik vektörünü temsil etmektedir. Her olabilirlik oranı  $P(f_i^k|c)/P(f_i^k|\bar{c})$ ,  $f_i^k$  vektörünün komşuluğundaki ilgili sınıf(lar)a ait özniteliklerin parametrik olmayan olasılık yoğunluk kestirimi yardımıyla hesaplanır.  $Z$  eğitim kümesindeki tüm süperpikselleri;  $B_i^k$  ise geri erişim kümesinde  $k$ . öznitelik  $f_i^k$ 'nin belli bir komşuluğundaki süperpikselleri temsil etsin. Dolayısıyla,

$$\frac{P(f_i^k|c)}{P(f_i^k|\bar{c})} = \frac{(n(c, B_i^k) + \varepsilon)/n(c, Z)}{(n(\bar{c}, B_i^k) + \varepsilon)/n(\bar{c}, Z)} = \frac{n(c, B_i^k) + \varepsilon}{n(\bar{c}, B_i^k) + \varepsilon} \times \frac{n(c, Z)}{n(\bar{c}, Z)} \quad (2.2)$$

Burada  $n(c, S)$ ,  $S$  kümesi içerisindeki  $c$  sınıfı etiketli süperpiksel sayısı,  $\varepsilon$  ise daha yumuşak olabilirlik kestirimi için küçük bir sabittir. Olabilirlik oranları hesaplandıktan sonra, test süperpikseli denklem (2.1)'de maksimum değeri veren sınıf ile etiketlenir. Bu yöntem Temel (Baseline) Algoritma olarak adlandırılmıştır.

### Bağlamsal Çıkarıma

Süperpikseller belli bir seviye uzamsal tutarlılık sağlamakla birlikte, etiketlemede bağlamsal kısıtlar tanımlayarak ayrıştırma doğruluğunu iyileştirmek mümkündür. Spesifik olarak görüntü etiketleme, süperpiksel etiketleri kümesi  $c = \{c_i\}$  üzerinde bir Markov Rasgele Alan (MRA) enerji enküçültme problemi olarak formüle edilebilir:

$$J(\mathbf{c}) = \sum_{s_i \in SP} D(s_i, c_i) + \sum_{(s_i, s_j) \in N} E(c_i, c_j) \quad (2.3)$$

Burada  $D$  ve  $E$  sırasıyla süperpiksel etiketlerinin veri maliyeti (data cost) ve düzlük maliyetidir (smoothness cost).  $SP$  görüntüdeki süperpiksel kümesi ve  $N$  ise bitişik süperpikselleri içeren bağlamsal komşuluktur. Veri terimi şöyle hesaplanır:  $D(s_i, c) = -w_i \sigma(L(s_i, c))$ . Burada  $L(s_i, c)$  log-olabilirlik oranı,  $w_i$  süperpiksel ağırlığı ve  $\sigma(\cdot)$  sigmoid fonksiyonudur.  $w_i$  ağırlığı süperpiksel boyutuyla orantılıdır:

$$w_i = \lambda |s_i| / \mu_s \quad (2.4)$$

Burada  $|s_i|$ ,  $s_i$ 'deki piksel sayısı,  $\mu_s$  ortalama süperpiksel boyutu ve  $\lambda$  düzlük sabitidir.

$E(c_i, c_j)$  düzlük maliyeti, etiketlerin birlikte görülme olasılıklarına göre belirlenir:

$$E(c_i, c_j) = -\log \left[ \left( P(c_i|c_j) + P(c_j|c_i) \right) / 2 \right] \times \delta(c_i \neq c_j) \quad (2.5)$$

Burada  $P(c|c_j)$  komşusu  $c_j$  etiketine sahip bir süperpikselin  $c$  etiketine sahip olma koşullu olasılığıdır. Bu olasılıklar eğitim setindeki komşu süperpiksellerin etiketlerinden kestirilir.



## 2.2 SuperParsing Performans Analizi

SuperParsing etiketleme doğruluğunu etkileyen pekçok farklı parametre değeri ve algoritma alt-adımları mevcuttur. Bunların bir kısmı (Tighe, 2013a) makalesinde incelenmiş ve tartışılmıştır. Projenin başlangıç aşamasında (Tighe, 2013a)'dan farklı olarak süperpiksel bölütleme sonuçlarının SuperParsing başarımı üzerindeki etkilerini analiz ettik.

Yöntemin başarımı 2,688 görüntü ve 33 etiket içeren SIFT Flow (Liu vd., 2008) veri kümesi üzerinde test edilmiştir. Deniz, sahil, dağ manzarası, cadde, bina silüeti gibi çeşitli sahneler içeren bu veri kümesinde *gökyüzü*, *deniz*, *ağaç*, *bina*, *araba* gibi farklı boyutlarda ve yoğunlukta 33 anlambilimsel sınıf yer almaktadır. Veri kümesi, 2,488 eğitim görüntüsü ve 200 test görüntüsü olmak üzere iki altkümeye ayrılmıştır. Benzetimlerde piksel-seviyesinde sınıflandırma doğruluğu (yani, doğru sınıflandırılan piksel yüzdesi) yanında sınıf-seviyesinde ortalama doğruluk yüzdelerinin karşılaştırılması yapılmıştır.

Literatürde pek çok süperpiksel bölütleme yöntemi mevcuttur. Bu yöntemler, ortaya çıkan süperpiksellerin boyutu, düzenliliği gibi model kısıtlarıyla birbirinden ayrılmaktadır. Süperpiksel tabanlı bölütlemelerde, görüntü görsel olarak anlamlı atomik bölgelere ayrılarak nesne kenarlarıyla uyumlu çok sayıda bölüt elde edilmektedir. Sınıflandırma performansı kullanılan süperpiksel algoritmasının özellikleri ve parametre ayarlarından önemli ölçüde etkilenmektedir.

Bu bölümde QuickShift (Vedaldi ve Soatto, 2008), çizge tabanlı bölütleme (ÇTB) ve SLIC (Achanta vd., 2012) yöntemleri test edilmiştir. Tablo 2.1'deki sonuçların da gösterdiği gibi, SuperParsing en yüksek piksel-seviyesinde etiketleme doğruluğuna ÇTB bölütleme yöntemi ile ulaşmaktadır. Bu nedenle projenin geri kalanında bu yöntem üzerinde odaklanılarak etiketleme doğruluğunun iyileştirilmesi için neler yapılabileceği irdelenmiştir.

Ayrıca Tablo 2.1'de ÇTB yönteminin parametre değerleri değiştirilerek performansın nasıl değiştiği de irdelenmiştir. ÇTB metodunda  $S$  parametresi en küçük süperpiksel boyutunu belirlerken  $K$  parametresi ise süperpiksellerin renk tutarlılığını kontrol etmektedir. Her iki parametrenin küçülmesi de süperpiksel boyutlarının küçülmesi sonucunu doğurmaktadır. Süperpiksel boyutlarının küçülmesi sonucunda sınıf seviyesinde ortalama doğruluğun bir miktar arttığı gözlenmiştir. Quickshift ile daha yüksek sınıf ortalaması elde edilmesinin temel nedeni de budur. Süperpiksel boyutu küçülüp toplam süperpiksel sayısı artınca eğitim setinde seyrek gözlenen "nadir" sınıflar (ör. *kapı*, *merdiven*) için daha fazla eğitim örneği ortaya çıkmaktadır. Eğitim örneğinin artması da bu sınıfların doğruluğunu yükseltmektedir. Quickshift yönteminin parametreleri değiştirilerek daha büyük süperpikseller elde edildiğinde piksel-seviyesinde doğruluk bir miktar artmakla birlikte ÇTB'nin performansını yakalayamamaktadır.

Tablo 2.1. Farklı bölütleme yöntemleri için SuperParsing ayrıştırma doğrulukları

Yöntem	Piksel-seviyesinde doğruluk (%)	Sınıf-seviyesinde doğruluk (%)
SLIC	71.6	24.1
QuickShift	71.1	32.0
ÇTB (K=200, S=100)	74.3	29.9
ÇTB (K=300, S=150)	74.1	30.2
ÇTB (K=400, S=200)	74.0	27.3
ÇTB (K=100, S=50)	73.4	30.2

Dolayısıyla SuperParsing algoritmasının başarımı incelendiğinde piksel-seviyesinde ortalama etiketleme doğruluğunun yüksekliğine karşın bazı “nadir” sınıflardaki doğru etiketleme yüzdelerinin oldukça düşük olduğu görülmektedir. Veri kümesinde temsili düşük olan *kapı*, *merdiven* gibi sınıflar için geri erişim kümesinde az sayıda süperpiksel tespit edilebilmekte; bu da bu sınıflar için hesaplanan olabilirlik oranlarının düşük çıkmasına neden olabilmektedir. Yukarıda belirttiğimiz gibi süperpiksel boyutlarının küçültülüp bu sınıflara atanan eğitim örneği miktarının artırılması mümkündür. Fakat piksel seviyesinde ortalama doğruluğu eniyilemek için daha baskın olan *gökyüzü*, *bina* gibi sınıflara en uygun süperpiksel boyutunun tercih edilmesi gerekmektedir. Dolayısıyla süperpiksel boyutlarının küçültülmesi veri kümesinde az bulunan ve boyutları küçük olan sınıflar için etiketleme doğruluğunu ciddi oranda arttırırken tüm pikseller için doğruluğun düşmesine neden olabilmektedir.

### 3. SUPERPARSING ALGORİTMASININ İYİLEŞTİRİLMESİ

Bu bölümde SuperParsing algoritmasının etiketleme başarımının iyileştirilmesine yönelik olarak, bölütleme, öznitelik seçimi ve sınıf olabirlik oranlarının hesabı için önerilen özgün yaklaşımlar açıklanmaktadır.

#### 3.1 Çok-hipotezli Süperpiksel Bölütleme ile Görüntü Ayrıştırma

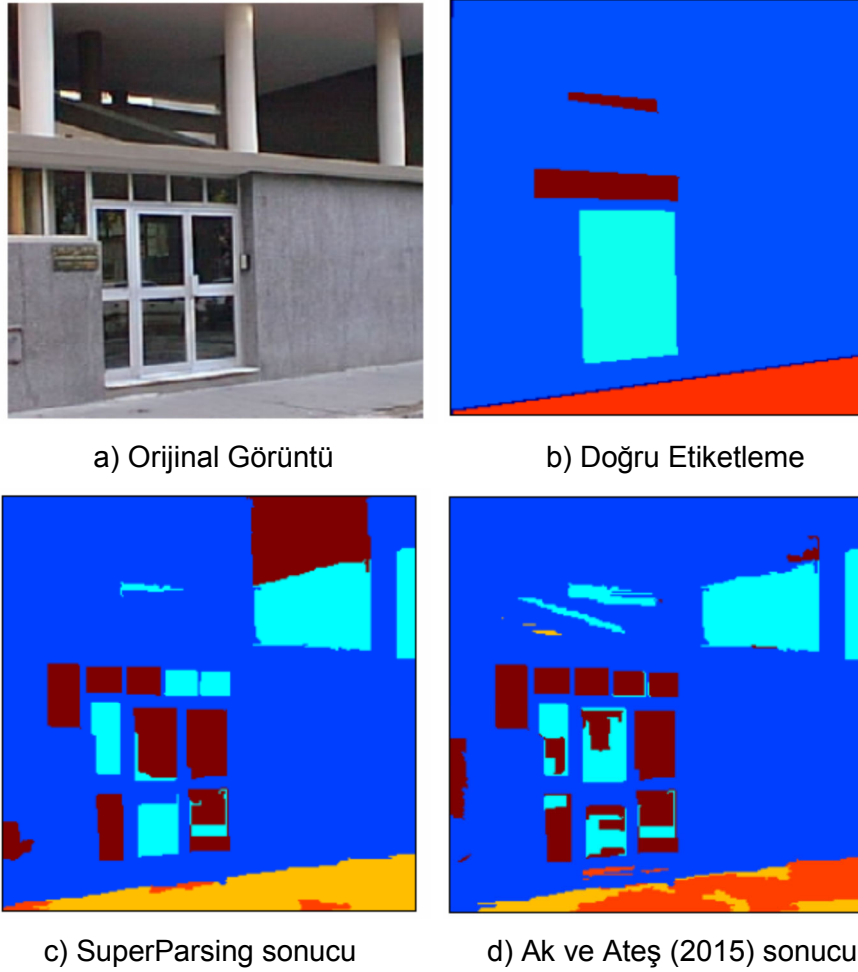
Bu proje için yürüttüğümüz ön çalışmada (Ak ve Ateş, 2015) SuperParsing algoritmasının alternatif süperpiksel gösterimleri için elde ettiği sonuçların kaynaştırılmasına dayanan basit bir yöntem test edilmişti. Bu bildiri sahne etiketleme doğruluğunu iyileştirmek için birden fazla süperpiksel bölütleme sonucunu sınıflandırıcı seviyesinde kaynaştıran bir yöntem önerilmiştir. Öncelikle SuperParsing algoritması kullanılarak süperpiksel etiketleri için log-olabirlik oranları tespit edilir. Daha sonra alternatif süperpiksel bölütleme senaryoları için hesaplanan olabirlik oranları piksel seviyesinde kaynaştırılarak, ilgili sahnenin bölütlenmesi ve etiketlenmesi tamamlanır.

Çok-hipotezli süperpiksel bölütleme olarak adlandırdığımız bu yaklaşımın motivasyonu, birden fazla bölütleme ile farklı sınıflar için uygun olacak alternatif gösterimlerin elde edilmesi ve böylece sınıflandırma doğruluğunun iyileştirilmesidir. Bölütleme alternatifi olarak, ÇTB'nin farklı parametre değerleriyle yürütülmesi yanında Quickshift bölütleme de uygulanmıştır. Birden fazla süperpiksel gösteriminin kullanılmasıyla, farklı nesne kategorileri için uygun bölütler tespit edilebilmiş, bu da daha doğru olabirlik oranlarının hesaplanmasını sağlamıştır. SuperParsing algoritmasının  $M$  farklı süperpiksel bölütleme alternatifi için çalıştırılması sonucu elde edilen log-olabirlik oranları  $L_m (1 \leq m \leq M)$  olsun. Tüm hipotezlerin birbirinden bağımsız olduğu varsayımı için, bir  $p$  pikselinin genel log-olabirlik oranı, ait olduğu süperpikselerin log-olabirlik oranlarının toplamına eşittir:

$$L(p, c) = \sum_{m=1}^M L_m(s_m, c | p \in s_m) \quad (3.1)$$

Burada  $p \in s_m$ ,  $p$  pikselinin  $m$  hipotezi altında  $s_m$  süperpiksel bölütüne ait olduğunu işaret etmektedir.  $L(p, c)$  log-olabirliğini eniyileyen  $c$  sınıfı,  $p$  pikselinin etiketi olarak atanır. Önerilen yöntem SIFT Flow veri kümesi üzerinde test edilmiş ve SuperParsing'den daha yüksek sınıflandırma doğruluğu elde edilmiştir (%74.3'e karşı %76.4). Eğitim seti üzerinde herhangi bir eniyileme ya da çapraz-sağlama (cross-validation) yapmadan sağlanan bu

performans artışı, daha akılcı bir modelleme ile çok ciddi iyileşme sağlanabileceğine işaretir. Şekil 3.1’de iki yöntemin sınıflandırma sonuçları görsel olarak karşılaştırılmıştır.



Şekil 3.1. Görsel karşılaştırma (mavi:bina, açık mavi:kapı, kahverengi:pencere, turuncu: kaldırım, sarı: yol)

Projenin ilk döneminde çok-hipotezli süperpiksel bölütlemenin performansının iyileştirilmesi üzerinde durulmuştur. Bunun için yapılan çalışmalar şöyle sıralanabilir:

- Nesne sınıfları için, alternatif süperpiksel bölütleri ve öznitelik betimleyicilerin koşullu olasılık dağılımları modellenmiştir:  $P(s_m, f^k | c)$  ( $s_m$  süperpiksel bölütünü,  $f^k$  ise  $k$ . betimleyiciyi temsil etmektedir). Bu modeller, eğitim kümesi üzerinde kernel yoğunluk kestirimi yaklaşımı ile belirlenmiştir (Bölüm 3.2).
- Alternatif bölütlerin eğitim kümesindeki performansı analiz edilmiştir. Bundan yola çıkarak bir test görüntüsü için en doğru bölütleme alternatifinin kestirimine çalışılmıştır (Bölüm 3.2.2).

### 3.2 SuperParsing’de Kernel-tabanlı Olabilirlik Kestirimi

Bu bölümde olabilirlik oranlarının ve dolayısıyla etiketleme doğruluğunun iyileştirilmesi için yerel kernel-tabanlı olasılık yoğunluk kestirimi kullanılması önerilmiştir (Ates vd., 2016). Her öznitelik vektörü için eniyi kernel bant-genişliğinin kullanılması ile öznitelik olasılık yoğunlukları, özellikle eğitim örneklerinin az olduğu durumlar için, daha doğru modellenenmiştir.

SuperParsing orta ve büyük veri kümelerinde başarılı sonuçlar sunmaktadır. Fakat özellikle eğitim örneklerinin az olduğu, düşük temsile sahip nadir sınıflar için kestirim doğruluğu düşmektedir. Sonuçta nadir sınıflara karşı negatif bir yanlılık (bias) söz konusudur ve bu da etiketleme doğruluğunu etkilemektedir.

Bu çalışmada sınıf-koşullu öznitelik olasılık yoğunlukları için kernel-tabanlı yoğunluk kestirimi kullanılmakta, ve böylece olabilirlik kestirimi ve genel ayırıştırma doğruluğu iyileştirilmektedir. Her öznitelik için Gauss kernellerinin bant-genişlikleri eğitim setinde eniyilendikten sonra test görüntülerinde yerel öznitelik yoğunluklarının daha doğru kestirimi için kullanılmaktadır. Kestirim öznitelik vektörlerinin uyarlanabilir yerel komşuluklarında yürütülmektedir. Komşuluklar her sınıfa ait eğitim örneklerinin miktarına göre belirlenir.

SuperParsing’deki veriye dayalı, parametrik olmayan olabilirlik kestiriminin iyi çalışması için veri kümesi yeterli büyüklükte olmalıdır ki test görüntüsündeki her bir süperpikselin öznitelik vektörlerinin etrafında yeterli sayıda yerel olarak kümelenmiş eğitim örnekleri bulunabilsin. Bu durumda her bir sınıftan eğitim örneklerinin miktarı sınıf-koşullu olasılığı için güvenilir bir ölçüm sağlar. Bununla birlikte nadir sınıflar için (ör. “kapı”, “merdiven”) özellikle dağılımın kuyruklarına doğru eğitim örneklerinin kümesi seyrek. Eğitim örneklerinin eğitim görüntülerinin daha küçük bir geri kazanım alt-kümesinden elde edilmesi problemi daha da artırmaktadır. Öznitelik komşuluğunun veya geri kazanım kümesinin büyüklüğünün artırılması, ilişkisiz örneklerle eşleşmenin artmasına ve sonuç olarak daha gürültülü olabilirlik kestirimlerine neden olmaktadır.

Bu çalışmada yerel öznitelik dağılımının daha doğru modellenmesi için kernel yoğunluk kestiriminin kullanılmasını önerdik. Öznitelik komşuluklarının büyüklüğü varolan eğitim örnekleri kümesine bağlı olarak uyarlanabilir şekilde belirlenir. Her bir örnek, bir Gauss kerneli ve örneğin test edilen öznitelik vektörüne mesafesi tarafından belirlenen bir ağırlığa göre yoğunluk kestirimine katkı sağlar. Dolayısıyla,  $f^k$  birbirinden bağımsız ve aynı dağılıma sahip olarak (i.i.d.) varsayarsak,  $P(f^k)$  olasılığı şu şekilde hesaplanabilir:

$$P(f^k) = \frac{1}{\Phi} \sum_{j=1}^{|Z|} \exp\left(-\frac{\|f^k - f_j^k\|^2}{\sigma_k^2}\right) \quad (3.2)$$

Burada  $|Z|$  eğitim kümesindeki süperpiksellerin sayısı,  $\sigma_k$  Gauss kernel bant-genişliği ve  $\Phi$  uygun bir düzgeleme çarpanıdır. Her bir öznitelik vektörü için, eniyi bant-genişliği  $\sigma_k$  sskernel methodu (Shimazaki ve Shinomoto, 2010) ile kestirilir.

Olabilirliklerin hesaplanmasında  $P(f^k|c)$  ve  $P(f^k|\bar{c})$ 'nin kestirimi için eniyi bant-genişlikleri kullanılır (Çoğu durumda  $P(f^k)$ ,  $P(f^k|c)$  ve  $P(f^k|\bar{c})$  için eniyi bant-genişlikleri aynıdır). Test süperpikseli  $s_i$  ve onun öznitelik vektörü  $f_i^k$  için, komşuluğundaki her bir eğitim örneği kernel kestirimi ile şu şekilde ağırlıklandırılır:

$$P(f_i^k|c) = \frac{1}{n(c,Z)} \sum_{j=1}^{n(c,B_i^k)} \exp\left(-\frac{\|f_i^k - f_{c,j}^k\|^2}{\sigma_k^2}\right) \quad (3.3)$$

Burada  $f_{c,j}^k$   $c$  sınıfındaki komşu vektörlerdir.  $P(f_i^k|\bar{c})$  benzer şekilde tanımlanır.

Olabilirlik oranlarının doğru kestirimi için  $B_i^k$  komşuluğunun boyutu kritiktir. Bir seçenek geri kazanım kümesindeki bütün örnekleri kullanmaktır. Fakat bu, fazla yumuşatılmış koşullu dağılım kestirimlerine neden olur. Onun yerine, test vektörü  $f_i^k$ 'nin en yakındaki  $R$  komşuyu, yani en düşük  $\|f_i^k - f_j^k\|$  uzaklığına sahip  $R$  örneği, kullanmayı öneriyoruz.

Bununla birlikte, özellikle test vektörü  $P(f^k|c)$  dağılımının kuyruklarındayken, en yakın  $R$  komşunun nadir sınıflardan örnek içermeye ihtimali düşüktür. Literatürde bu problem için önerilen çözüm nadir sınıflar için fazladan eğitim örnekleri kullanarak daha dengeli eğitim kümesi tasarlamaktır. Bu çözüm sınıf-seviyesinde ortalama doğruluğu iyileştirir fakat piksel-seviyesi doğruluğa fayda etmez. Bu çalışmada, genel piksel-seviyesindeki doğruluğu iyileştirmek amacıyla, her bir test öznitelik vektörü  $f_i^k$  için  $B_i^k$  komşuluğunu uyarlanabilir şekilde seçmeyi öneriyoruz. Prosedür şu şekildedir:  $B_i^k$  kümesindeki en yakın  $R$  komşunun etiketlerini kontrol ediyoruz; eğer bütün örnekler aynı sınıftan veya 2 farklı sınıftan geliyorsa,  $r$  tane yeni  $c$  sınıfından en düşük  $\|f_i^k - f_{c,j}^k\|$  uzaklığına sahip birer örnek,  $f_{c,j}^k$ ,  $B_i^k$  kümesine eklenir. Böylelikle sınıflar yeterli temsil edilmediğinde, komşuluk kümesi uyarlamalı olarak genişletilir.

### 3.2.1 Benzetimler

Benzetimlerde kernel-tabanlı kestirim yöntemi orijinal SuperParsing ve onun modifiye versiyonları ile karşılaştırılmıştır. Algoritmaların başarımı SIFT Flow veri kümesi üzerinde test edilmiştir. Benzetimlerde piksel-seviyesinde sınıflandırma doğruluğu (yani, doğru

sınıflandırılan piksel yüzdesi) yanında farklı sınıflar için doğruluk yüzdelerinin karşılaştırılması yapılmıştır.

Algoritmanın değişkenleri şu şekilde ayarlanmıştır. SuperParsing yöntemindeki gibi geri kazanım kümesinin büyüklüğü 200 görüntüdür. SuperParsing’de  $K=200$ ,  $S=100$ . En yakın komşu kümesi büyüklüğü  $R$ , geri kazanım kümesindeki örneklerin ortalama sayısı ile orantılıdır. 3 farklı alternatifi test ediyoruz:  $(K=300, S=150, R=20)$ ,  $(K=200, S=100, R=30)$ ,  $(K=400, S=200, R=15)$ . Ek sınıf sayısı  $r=4$  olarak sabitlendi.

Ayrıca SuperParsing’in 2 farklı modifikasyonu ile karşılaştırma yaptık: Uyarlanabilen komşu kümesi (Eigen ve Fergus, 2012) ve Çok-hipotezli SuperParsing (Ak ve Ateş, 2015). Uyarlanabilen komşu kümeler yaklaşımında,  $B_i^k$  kümesindeki her bir eğitim örneğine belli bir ağırlık verilir. Ağırlıklar eğitim kümesinde bayır inişi tabanlı eniyileme ile belirlenir.

Tablo 3.1 test edilmiş yöntemlerin piksel seviyesindeki etiketleme doğruluğunu vermekte. İlk sütun süperpiksellerin herhangi bir bağlamsal modellenmesi olmadan sadece Temel Algoritma sonuçları için. İkinci sütundaki sonuçlar bağlamsal çıkarım için MRA modeli (Tighe, 2013a) kullanılmasıyla elde edilmiştir. Ayrıca önerilen kernel-tabanlı metot çok-hipotezli yaklaşımın içine dahil edilmiş ve tabloda Kernel+Çok-hipotezli olarak adlandırılmıştır (Çok-hipotezli yaklaşımın MRA modeli Bölüm 4’te ele alınacaktır). Parametre ayarları şu şekildedir: SuperParsing1 ( $K=200, S=100$ ), SuperParsing2 ( $K=300, S=150$ ), Kernel-tabanlı ( $K=300, S=150$ ), ve Kernel+Çok-hipotezli ( $(K=300, S=150)$  ve  $(K=400, S=200)$  kullanarak iki hipotez).

Tablo 3.1. SIFT Flow veri kümesi için piksel etiketleme doğruluğu

Metot	Doğruluk Yüzdesi	
	Temel Algoritma	MRA
SuperParsing1	74.3	76.2
SuperParsing2	74.1	75.8
Uyarlanabilen komşu kümesi	76.8	77.1
Çok-hipotezli	76.4	-
Kernel-tabanlı	76.6	77.5
Kernel+Çok-hipotezli	77.8	-

Kernel-tabanlı yöntem SuperParsing’e göre Temel’de %2’den fazla ve MRA kullanıldığında yaklaşık %1.3 oranında iyileşme sağlamaktadır. MRA modeli ile birlikte yöntem, uyarlanabilen komşu kümesi (Eigen ve Fergus, 2012) yaklaşımından daha iyi bir performans sergilemektedir. Her iki yöntemde de eğitim örnekleri ağırlıklarındırılmaktadır. Bununla birlikte, Eigen ve Fergus (2012) betimleyici ağırlıklarını eğitim süreci sonunda sabitler; önerilen yöntemdeki ağırlıklar ise öznitelik vektörleri arasındaki mesafelere bağlı

olarak her bir test süperpikselli için uyarlanabilen şekilde belirlenir. Kernel-tabanlı kestirim için gerekli bant-genişliği eniyilemesi Eigen ve Fergus (2012)'de kullanılan ağırlık adaptasyonundan daha basit ve hızlıdır. Ayrıca Eigen ve Fergus (2012) etiketleme doğruluğunu iyileştirmek için eğitim kümesinde ters çevrilmiş görüntüler ve nadir sınıflar için fazladan geri kazanım kümesi örnekleri kullanmaktadır.

Çok-hipotezli olabilirlik hesabı, kernel-tabanlı yöntem ile birleştirildiğinde piksel sınıflandırma doğruluğu %77.8'e ulaşmaktadır. Çok-hipotezli yapıda MRA bağlamsal modelleme kullanıldığında doğruluk oranının daha da artacağını Bölüm 4'te göreceğiz.

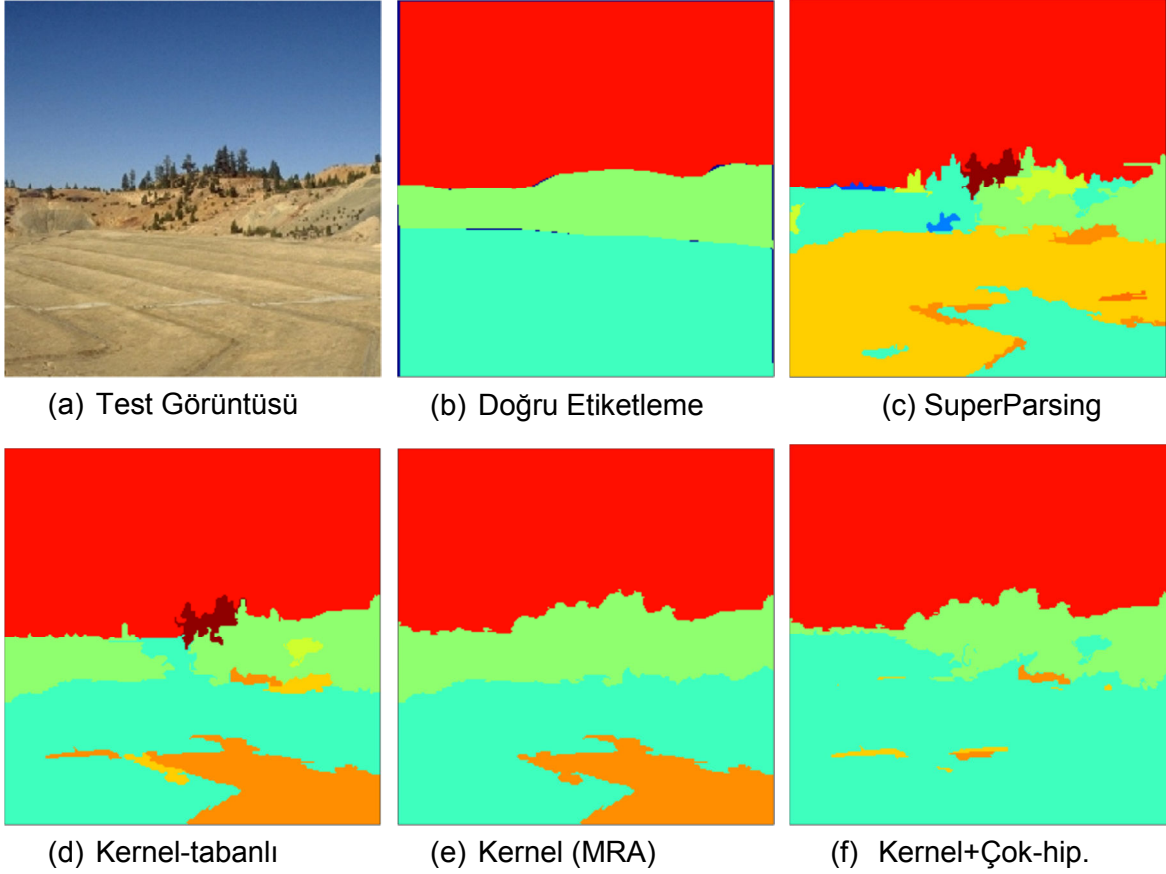
Tablo 3.2'de bazı sınıfların doğruluk yüzdeleri Kernel+Çok-hipotezli ve SuperParsing yöntemleri için karşılaştırılmıştır. Tablodan görüldüğü üzere, Kernel+Çok-hipotezli kestirim bazı orta-boyutlu ve küçük-boyutlu sınıfların, örneğin *çimen* ve *kapı*, etiketleme doğruluklarını geliştirir. Diğer taraftan, *arazi*, *bina*, *dağ* gibi büyük-boyutlu sınıflarda da piksel-seviyesi etiketleme doğruluğu artmış ve sonuç olarak genel piksel-seviyesi etiketleme doğruluğu yükselmiştir.

Tablo 3.2. Farklı sınıflar için sınıflandırma doğrulukları

Sınıf	SuperParsing1	Kernel + Çok-hipotezli
<i>yaya geçidi</i>	<b>23.9</b>	14.7
<i>merdiven</i>	12.1	<b>16.1</b>
<i>kapı</i>	10.1	<b>17.4</b>
<i>çimen</i>	35.2	<b>46.4</b>
<i>araba</i>	<b>51.3</b>	45.4
<i>arazi</i>	47.4	<b>70.5</b>
<i>dağ</i>	67.0	<b>73.7</b>
<i>bina</i>	84.4	<b>89.0</b>
<i>gökyüzü</i>	91.2	<b>91.9</b>

Şekil 3.2'de test edilmiş yöntemlerin etiketleme doğrulukları görsel olarak karşılaştırılmıştır. SuperParsing yöntemi *dağ* ile *arazi* ve *arazi* ile *yol* sınıflarını karıştırmaktadır. Kernel-tabanlı yöntemde ise daha az görsel hata var, fakat *arazi* sınıfının bazı kısımları ile *kum* sınıfı karıştırılmış. MRA+Kernel ve Kernel+Çok-hipotezli yöntemleri kernel-tabanlı yöntemin sonuçlarını daha da iyileştirmektedir.





Şekil 3.2. Görsel karşılaştırma (kırmızı: gökyüzü, yeşil: dağ, açık yeşil: arazi, açık turuncu: yol, turuncu: kum)

### 3.2.2 Çok-hipotezli SuperParsing’de Ağırlıklı Olabilirlik Kestirimi

Çok-hipotezli SuperParsing’de alternatif bölütler için hesaplanan olabilirlik oranları toplanarak daha doğru bir kestirim yapılabildiği gösterilmiştir. Bu bölümde çok-hipotezli SuperParsing’in performansının iyileştirilmesi için olabilirlik oranlarının ağırlıklı toplanması üzerine bir yaklaşım sunulmuştur. Bu ağırlıklar her test görüntüsü için uyarlanabilir ve eğitim verisi üzerinde yürütülen analizlerle belirlenmiştir.

Alternatif bölütlerin ağırlıklandırılmasında kriter olarak eğitim verisindeki görüntülerde gösterdikleri performans kullanılmıştır. Bir test görüntüsünün geri erişim kümesi  $\{\Gamma_t\}_{1 \leq t \leq 200}$  ve bu kümedeki görüntüler için  $m$  hipotezinin hesaplanan doğrulukları  $a_t^m$  olsun. Geri erişim kümesinde test görüntüsüne benzer sahnelerin bulunduğu varsayımı altında  $a_t^m$  değerleri ile test görüntüsündeki beklenen doğrulukların ilişkili olduğu değerlendirilmiştir. Dolayısıyla hipotezlerin ağırlıklandırılmasını şu şekilde yapmayı öneriyoruz:

$$L(p, c) = \sum_{m=1}^M w_m L_m(s_m, c | p \in s_m) \quad (3.4)$$

Ağırlıklar şöyle belirlenir:

$$w_m = \frac{1}{200} \sum_{t=1}^{200} 1(a_t^m = \max\{a_t^\alpha\}_{1 \leq \alpha \leq M}) \quad (3.5)$$

Burada  $1(\cdot)$  gösterge fonksiyonudur. Dolayısıyla bir test görüntüsü için her hipotezin log-olabilirlik değeri, geri erişim kümesinde en yüksek doğruluğu sağladığı görüntü sayısı oranında ağırlıklandırılır. Bu yaklaşımın Kernel+Çok-hipotezli yönteminde kullanılmasıyla etiketleme doğruluğu %77.8'den %78.0'a çıkmış ve %0.2'lik bir artış sağlanmıştır.

Benzer bir yaklaşım süperpiksel/piksel seviyesinde ağırlık uyarlaması için de kullanılabilir. Bu kez her test süperpikseli  $s_i$  için  $B_i^k$  komşuluklarındaki ( $1 \leq k \leq 20$ ) tüm eğitim örneklerinde farklı hipotezlerin etiketleme doğruluğu dikkate alınabilir. Farklı sınıflandırıcıların öznitelik uzayında yerel doğruluk performansına bakarak sınıflandırıcıların kaynaştırılması ile ilgili literatürde de benzer çalışmalar vardır ((Woods vd., 1996), (Duin, 2002)). Belli bir öznitelik vektörünün komşuluğundaki eğitim örnekleri için farklı sınıflandırıcıların gösterdiği sınıflandırma doğrulukları, o öznitelik vektörüne sahip test örneklerinde sınıflandırıcı kaynaştırma için kullanılmıştır. Fakat alternatif süperpiksel bölütleme ile elde edilen eğitim örnek uzayları birbirinden farklıdır ve farklı hipotezlerin kendi öznitelik uzaylarındaki yerel doğruluklarının birbiriyle karşılaştırılmasını zorlaştırmaktadır.

Farklı ayıştırma sonuçlarının uyarlanabilir birleştirilmesine yönelik olarak geliştirdiğimiz çok-hipotezli MRA modeli Bölüm 4'te detaylı şekilde ele alınacaktır. Bu esnek ve uyarlanabilir yöntem yerel seviyede süperpiksellerin komşuluk ilişkilerini kullanarak sınıf olabilirlik oranlarını ve etiketleme alternatiflerini eniyi biçimde kaynaştırabilmektedir.

### 3.3 Süperpiksel Bölütleme ve Görüntü Ayırıştırma Ortak Eniyileme

Bu bölümde sunulan çalışmaların amacı, ayırıştırma başarımını iyileştirilebilmek adına, sınıflandırma sonuçlarını etkileyen bütün algoritma bileşenlerinin daha uyarlanabilir şekilde seçimi için yöntemler geliştirilmesidir. Uyarlama görüntü seviyesinde ya da her nesne sınıfı için olabileceği gibi her test süperpikselinin sınıflandırılmasında eniyi parametrelerin belirlenmesine de çalışılmıştır. Sonuçta her test görüntüsü/süperpikseli ve her nesne sınıfı için eniyi bölütleme kriterleri, öznitelik betimleyicileri ve eğitim örneklerinin tahmini amaçlanmıştır. Bu amaç doğrultusunda yapılan çalışmalar şu şekilde sıralanabilir:

- Eğitim kümesi üzerinde nesne sınıflarının ve bölütlerinin analizi: öznitelik vektörlerinin vektör uzayında sınıf içi ve sınıflar arası dağılımları incelenmiştir. Süperpiksel bölütlerinin öznitelik vektörlerinin nesne sınıflarını ayırıştırma kapasitelerinin ölçümü için farklı metrikler önerilmiş ve analiz edilmiştir.
- Çeşitli metriklere göre bir nesne sınıfını diğerlerinden eniyi ayırt eden öznitelikler tespit edilmiştir. Bu metrikler alternatif süperpiksel bölütlerinin karşılaştırmalı analizi için de kullanılmıştır.
- Her test süperpikseli için, komşuluğundaki eğitim örnekleri ve bunların özniteliklerinin yerel dağılımına bağlı olarak uyarlanabilir bir sınıf-koşullu olasılık dağılım kestirimi yöntemi geliştirilmiştir.

Aşağıdaki bölümlerde yapılan çalışmalar daha ayrıntılı olarak ele alınmıştır.

### Süperpiksel bölüt ve özniteliklerinin analizi

İki-hipotezli SuperParsing'de bir piksel ait olduğu iki bölütten gelen toplam 40 farklı öznitelik vektörü için hesaplanan sınıf-koşullu log-olabilirlik oranlarının toplamına göre sınıflandırılmaktadır. Tablo 3.3'te çeşitli öznitelik vektörlerinin tek başlarına ve birlikte kullanılmaları durumunda piksel-seviyesinde ve farklı sınıflar için elde edilen etiketleme doğrulukları iki farklı bölütleme ( $(K=300, S=150)$  ve  $(K=400, S=200)$ ) için listelenmiştir.

Tablo 3.3'ten anlaşılacağı gibi özniteliklerin sınıflandırma başarımları farklılıklar göstermekle birlikte birbirlerini tamamlayıcı olmaları sayesinde birlikte kullanıldıklarında etiketleme doğruluğuna hepsi katkıda bulunmaktadır. Bu tablodan çıkan bir diğer sonuç ise, bir öznitelik vektörünün başarımında süperpikselin sınıfı ve bölütleme yöntemine göre kayda değer değişiklikler olabildiğidir. Örneğin *top\_height* özniteliği *kum* sınıfı için *yol* sınıfına göre çok daha başarılıdır; çünkü *top\_height* süperpikselin görüntü içindeki yüksekliğini göstermektedir ve sahnede kum bölgeleri gökyüzü ve denizin altında kaldığı için *kum* süperpiksellerinin yüksekliği düşüktür. Halbuki *yol* süperpikselleri sahnede farklı bölgelerde yer alabileceği için yükseklik bilgisi ayırt edici değildir.

Tablo 3.3. Farklı öznitelikler için sınıflandırma doğrulukları

Öznitelik	$(K=300, S=150)$					$(K=400, S=200)$				
	piksel	dağ	bina	yol	kum	piksel	dağ	bina	yol	kum
<i>sift_hist_int</i>	51.9	32.3	38.2	63.0	49.4	53.8	30.9	46.4	62.6	41.8
<i>int_text</i>	36.3	14.5	15.6	46.9	23.7	37.0	14.0	18.7	50.8	26.4
<i>bb_extent</i>	31.3	10.9	25.7	43.9	55.5	28.4	12.2	26.9	44.2	26.7
<i>top_height</i>	21.8	10.6	7.4	2.4	33.5	24.9	7.6	9.7	6.1	20.5
<i>color_hist</i>	33.1	13.8	12.1	33.9	13.3	33.0	15.0	12.2	33.5	8.1
<i>gist_int</i>	40.8	14.1	27.2	58.8	42.3	43.2	15.0	30.2	58.7	17.6
hepsi	76.6	72.6	88.5	80.5	15.8	76.4	72.1	88.7	83.4	12.4

Bölüt alternatiflerinin çeşitli sınıflarda gösterdikleri performans farklarından faydalanabilmek için basit bir uyarlamalı yöntemi test ettik. İki hipotezli yöntem için öncelikle her iki alternatifin etiketleme doğruluğu eğitim kümesi üzerinde karşılaştırılır ve her sınıf için en yüksek etiketleme doğruluğu veren hipotez belirlenir. Bu bilgiyi  $h(c) \in \{0,1,2\}$  ile temsil edelim. Eğer iki hipotezin doğrulukları arasında kayda değer bir fark yoksa o zaman  $h(c) = 0$ . Bir test pikselinin ait olduğu iki süperpikselin etiketleri  $c_1, c_2$  ve çok-hipotezli yöntemden gelen etiket  $c_j$  olsun. Bu durumda etiket tahmini  $c_f$ 'yi şu şekilde güncellemeyi öneriyoruz:

Eğer  $h(c_j) = 1$  ise  $c_f = c_1$

Eğer  $h(c_j) = 2$  ise  $c_f = c_2$

Aksi takdirde  $c_f = c_j$

Genelde  $h(c) \neq 0$  olan sınıflar etiketleme doğruluğu düşük nadir sınıflara denk gelmektedir. Dolayısıyla etiket güncellemesinin genel piksel-seviyesindeki başarımlar üzerinde etkisi sınırlı kalmaktadır. Farklı hipotezler için yapılan benzetimlerde etiketleme doğruluğunda %0.2 - %0.5 arası bir iyileşme gözlenmiş ve maksimum doğruluk olarak %78 elde edilmiştir.

### Öznitelikler ve sınıflar arasındaki karşılıklı bilginin kestirimi

Tablo 3.3 ilginç olmakla birlikte özniteliklerin ortak başarımına ışık tutmamaktadır. Başka deyişle tümü birlikte kullanıldığında hangi özniteliğin hangi durumda ne kadar katkı sağladığının analizi çok daha karmaşıktır. Bu bölümde bu katkıyı ölçebilmek için bir metrik önerilmiş ve test edilmiştir. Burada hedeflenen, uygun metriklerin kullanılmasıyla her test görüntüsü/süperpikseli için eniyi bölüt ve özniteliklerin tahmin edilmesi ve böylece daha doğru log-olabilirlik oranlarının hesaplanmasıdır.

Bir özniteliğin sınıfları birbirinden ayırt etme gücünün ölçümü için sınıfların önsel dağılımı  $P(c)$  ve öznitelik-koşullu sınıf dağılımı  $P(c|f^k)$  karşılaştırılabilir. Eğer  $c$  ve  $f^k$  birbirinden bağımsız ise  $P(c) = P(c|f^k)$ . Bu durumda karşılıklı bilgi miktarı (mutual information)  $I(c, f^k) = 0$ .

$$I(c, f^k) = \int \sum_j P(c_j, f^k) \log \frac{P(c_j, f^k)}{P(c_j)P(f^k)} df^k \quad (3.6)$$

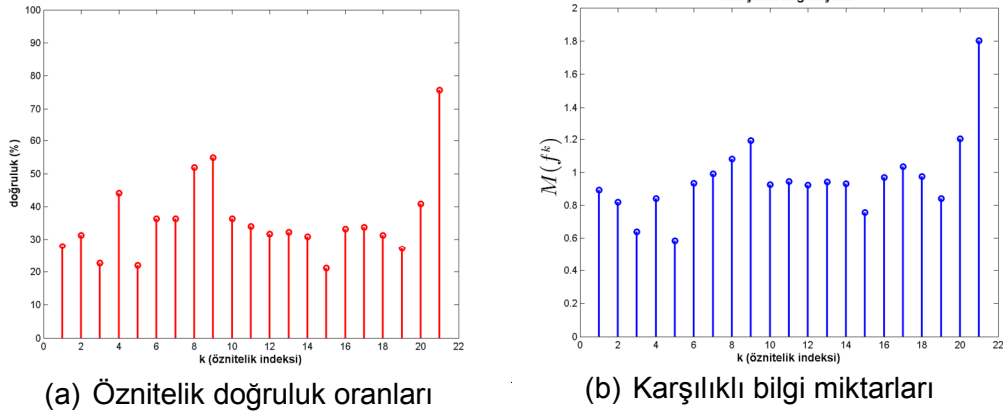
Dolayısıyla  $I(c, f^k)$  ne kadar büyükse özniteliğin o kadar ayırt edici olduğu söylenebilir. Bir süperpiksel  $s_i$  için komşuluğundaki eğitim örneklerinin yukarıdaki ifadeye katkısı şöyle hesaplanabilir:

$$I_{s_i}(c, f^k) = n(c, B_i^k) \log \left( \frac{n(c, B_i^k) |Z|}{n(c, Z) |B_i^k|} \right) \quad (3.7)$$

Bu metriği tüm sınıflar ve tüm test süperpikselleri üzerinden toplarsak:

$$M(f^k) = \frac{1}{\Phi} \sum_{s_i} \sum_j I_{s_i}(c_j, f^k) \quad (3.8)$$

Burada  $\Phi$  toplamdaki tüm örneklerin sayısıdır. Şekil 3.3'te farklı öznitelik vektörleri ( $1 \leq k \leq 20$ ) için toplam karşılıklı bilgi miktarları  $M(f^k)$  ile test görüntülerindeki etiketleme doğrulukları arasındaki ilişki gösterilmiştir. Görüldüğü gibi karşılıklı bilgi miktarlarıyla doğruluk oranları arasında güçlü bir ilinti vardır ve karşılıklı bilgi miktarı arttıkça doğruluk oranı genelde artmaktadır. Bu grafikte  $k = 21$ , tüm öznitelik vektörlerinin birlikte kullanıldığı durumu temsil etmektedir. Bu durumda  $M(\{f^l\}_{1 \leq l \leq 20})$  hesabı için toplam log-olabilirlik oranları  $L(s_i, c)$  kullanılarak  $P(c_j | \{f^l\}_{1 \leq l \leq 20})$  ve  $P(c_j, \{f^l\}_{1 \leq l \leq 20})$  olasılıkları bulunmuştur. Grafikten de anlaşıldığı gibi karşılıklı bilgi metriği tüm özniteliklerin birlikte kullanılmasıyla elde edilen doğruluk artışını hassas bir şekilde göstermektedir. Her iki grafik arasındaki ilinti katsayısı 0.90 gibi yüksek bir değerdir.



Şekil 3.3. Farklı öznitelikler için etiketleme doğruluğu – karşılıklı bilgi miktarları karşılaştırması (( $K=300$ ,  $S=150$ ) bölütleme için).

Karşılıklı bilgi miktarının tüm test kümesinde sınıflandırma doğruluğunun gurbüz bir ölçütü olduğu görülmektedir. Bu gözlemden yola çıkarak, daha yerel seviyede her test görüntüsü ya da süperpikseli için, bu ölçütü alternatif öznitelik ve bölütlerin sınıflandırma başarımını kestirmede kullanmayı denedik. Bu amaçla denklem (3.7)'deki değer farklı sınıflar üzerinden toplamı,  $\frac{1}{|B_i^k|} \sum_j I_{s_i}(c_j, f^k)$ , bir süperpiksel için farklı öznitelik vektörlerinin

doğru sınıflandırma kapasitesinin ölçütü olarak test edildi. Geliştirilen bu ölçüt, görüntü/ süperpiksel seviyesinde alternatif bölüt ve öznitelikler arasında tercih yapmak için ya da alternatiflere ağırlık atamak için kullanıldı. Fakat yapılan testler elde edilen performans artışlarının kısıtlı kaldığını (%0.2'den az) göstermiştir. Süperpiksel seviyesinde uyarlanabilir bir yöntem geliştirmek için her süperpikselin komşuluğundaki öznitelik vektörlerin dağılımı ile ilgili daha gelişmiş bir modele ihtiyaç olduğu sonucuna varılmıştır.

### Koşullu olasılıkların kestiriminde yumuşak ağırlıklandırma

Uyarlanabilir bir modelleme ile hedeflenen temel amaç sınıf-koşullu olasılık dağılımlarının daha doğru kestirilmesidir. Tüm öğrenme problemlerinde olduğu gibi, eğitim kümesinden öğrenilmeye çalışılan  $P(f^k|c)$  koşullu olasılıklarının kestirimindeki hatanın iki temel bileşeni model yanlılığı (bias) ve değişintisidir (variance). Bu iki hata bileşeni arasında doğru bir ödünleşim ile daha başarılı öğrenme algoritmaları geliştirilmesinin mümkün olduğu düşünülmektedir.

Bir  $s_i$  süperpikselin komşuluğundaki her eğitim örneği  $f_{c^*,j}^k$ ,  $P(c|f^k)$  için çok basit bir kestirim olarak görülebilir:

$$\hat{P}_j(c|f^k) = P(c|f_{c^*,j}^k) = \delta(c = c^*) \quad (3.9)$$

Burada  $\delta(\cdot)$  delta (birim dürtü) fonksiyonudur. Bu kestirimin yanlılığı düşük ama değişintisi yüksektir. Komşuluktaki tüm örnekler kullanıldığında:

$$\hat{P}(c|f^k) = \frac{1}{|B_i^k|} \sum_{j=1}^{|B_i^k|} \hat{P}_j(c|f^k) = \frac{1}{|B_i^k|} \sum_{j=1}^{|B_i^k|} P(c|f_{c^*,j}^k) \quad (3.10)$$

Böylelikle değişinti düşürülerek daha gürbüz bir kestirim elde edilmeye çalışılmaktadır. Kestirim hatası üç bileşenine ayrılarak incelenebilir:

$$E[(P - \hat{P})^2] = \text{Bias}[\hat{P}]^2 + \text{Var}[\hat{P}] + \sigma^2 \quad (3.11)$$

Burada  $\sigma^2$  indirgenemez hatayı temsil etmektedir. Yanlılık hatası:

$$\text{Bias}[\hat{P}] = E[\hat{P}] - P \quad (3.12)$$

Ve değişinti hatası:

$$\text{Var}[\hat{P}] = E[(\hat{P} - E[\hat{P}])^2] \quad (3.13)$$

Yanlılık ve deęişinti arasında daha iyi bir ödünleşim sağlayabilmek için  $\hat{P}_j(c|f^k)$  kestirimi olarak bir birim dürtü kullanmak yerine sınıf olasılıklarına daha yumuşak bir atama yapmayı öneriyoruz.  $P(c|f_{c^*,j}^k)$  olasılıklarının nasıl seçileceęi ise eğitim kümesinde  $f_{c^*,j}^k$  örneğinin kendi komşuluğundaki örneklerin olasılık dağılımları ile belirlenecektir. Dolayısıyla tüm eğitim kümesi örnekleri üzerinde döngülü şekilde  $P(c|f_{c^*,j}^k)$  koşullu olasılıkları aşğıdaki algoritma ile belirlenir.

$1 \leq j \leq |Z|$  için:

**İklendirme:**  $P(c|f_{c^*,j}^k) = \delta(c = c^*)$

**Güncelleme:** Koşullu olasılıklar her örneğin komşuluğundaki örnekler kullanılarak güncellenir:

$$P(c|f_{c^*,j}^k) = (1 - \alpha)P(c|f_{c^*,j}^k) + \frac{\alpha}{|B_j^k|} \sum_{l=1}^{|B_j^k|} \hat{P}_l(c|f_{c^*,j}^k) \quad (3.14)$$

Burada  $\alpha$  güncelleme miktarını temsil etmektedir ve eniyi yanlılık-deęişinti ödünleşimini sağlayacak şekilde seçilmelidir.

Güncelleme yeterli iterasyondan sonra sonlandırılır. Eğitim kümesinden elde edilen  $P(c|f_{c^*,j}^k)$  olasılıkları test kümesindeki süperpikseller için denklem (3.14)'deki gibi  $\hat{P}(c|f^k)$  ve log-olabilirlik oranlarının kestiriminde kullanılır. Önerilen yöntemde en kritik parametre  $\alpha$  güncelleme katsayısıdır.

Farklı  $\alpha$  deęerleri ve iterasyon sayılarıyla yapılan testlerde, sınıf ortalama doğruluğunda %1'e varan artışlar gözlenirken piksel-seviyesinde doğruluk artışı %0.2'nin altında kalmıştır. Dolayısıyla komşuluktaki örneklerin daha yumuşak ağırlıklandırılması nadir sınıflar için daha başarılı etiketleme sağlamakla birlikte bu sonuç piksel seviyesinde etiketleme doğruluğuna fazla yansımamaktadır.

### 3.4 Süperpiksel Öznitelikleri için Kodlama Yöntemleri

Önceki bölümde süperpiksel özniteliklerinin ayrıştırma başarımları kapsamlı şekilde ve farklı metriklerle analiz edilmişti. Süperpiksel seviyesinde uyarlanabilir öznitelik seçimi ve ağırlıklandırma ile etiketleme doğruluğunda kayda deęer bir artış elde edilemedięi gözlenmiştir.

Özniteliklerin sınıflandırma başarımını iyileştirmek için literatürde önerilen çok sayıda öznitelik kodlama yöntemi mevcuttur. Bu bölümde iki farklı kodlama yönteminin SuperParsing etiketleme doğruluğu üzerindeki etkisi kapsamlı benzetimlerle analiz edilmiştir. Elde edilen sonuçlardan bir konferans bildirisi yayınlanmıştır (Ateş ve Süneci, 2017a).

Süperpiksel seviyesinde öznitelikler genelde kelime torbası (bag of words) yaklaşımı ile tespit edilmektedir. Öncelikle süperpiksele ait her piksel için öznitelik vektörü hesaplanır. Daha sonra bu öznitelikler vektör nicemlenir ve nicemleme indislerinin histogramı çıkarılır. Hesaplanan histogram süperpiksel için öznitelik betimleyici olarak kullanılır.

Son yıllarda kelime torbası modelini iyileştirmek için, sabit nicemleme kullanan bir kodlama yerine, öznitelik vektörleri ile ilgili daha fazla bilginin korunmasını sağlayan özgün kodlama yöntemleri önerilmiştir (Chatfield vd., 2012; Huang vd., 2014). Bu yöntemler sahne ve nesne sınıflandırma problemlerinde test edilmiş ve sınıflandırma başarımında önemli artışlar sağladıkları gösterilmiştir.

Literatürde süperpiksel özniteliklerine farklı kodlama yöntemleri uygulayan görüntü ayrıştırma çalışmaları pek yoktur. George (2015)'te SIFT öznitelikleri için Fisher vektör kodlama kullanılmıştır. Aşağıda detaylandırılan çalışmada süperpiksel SIFT öznitelikleri için Kernel Kod-tablosu Kodlama (KKT) (van Gemert vd., 2008) ve Yerellik-Kısıtlı Doğrusal Kodlama (YDK) (Wang vd., 2010) yöntemleri test edilmiştir.

### 3.4.1 Öznitelik Kodlama Yöntemleri

Sınıflandırma problemlerinde üç temel adımdan bahsedilebilir:

- i) yerel (piksel-seviyesinde) özniteliklerin çıkarılması;
- ii) öznitelik vektörlerinin kodlanması ve birleştirilmesi (pooling) ile görüntü/nesne/süperpiksel için öznitelik betimleyici elde edilmesi;
- iii) görüntü/nesne/süperpiksel betimleyicinin sınıflandırılması.

Kelime torbası modelindeki temel yaklaşım, öznitelik vektörlerinin bir kod-tablosundaki en yakın kodsözcüğüne atanarak nicemlenmesi ve nicemlenmiş özniteliklerin histogramının betimleyici olarak kullanılmasıdır. Kod-tablosundaki kodsözcükleri (yani, görsel kelimeler) genelde eğitim setindeki vektörlerde *kmeans* gibi bir topaklandırma algoritması kullanılarak tespit edilir.

Literatürde son dönemde bu sabit nicemleme yöntemine alternatif olarak önerilen pek çok öznitelik kodlama yaklaşımı sunulmuştur. Bunlardan bazıları: Yerellik-Kısıtlı Doğrusal Kodlama (Wang vd., 2010), Fisher Vektör Kodlama (Perronnin vd., 2010), Süper Vektör Kodlama (Zhou vd., 2010), ve Kernel Kod-tablosu Kodlama (KKT) (van Gemert vd., 2008). Sabit nicemleme/eşleme yerine bu gelişmiş kodlama yöntemlerinin kullanılması,



nicemlemeden kaynaklanan hatanın ve bilgi kaybının azaltılarak daha ayrıştırıcı betimleyiciler elde edilmesini ve böylece sınıflandırma başarımının artmasını sağlamıştır.

### Kernel Kod-tablosu Kodlama (KKT)

Sabit nicemleme/eşleme yöntemi, verilen sözlükteki en yakın görsel kelime ile her bir betimleyici vektörü eşleştirir. İki birbirine benzemeyen öznitelik vektörü herhangi bir ayırım yapılmaksızın aynı görsel kelimeye atanabilir. Kodsözcüğü anlam belirsizliği ile ilgilenmemesinden dolayı önemli bilgi kaybına neden olur. Bu yöntemde betimleyici, sözlükteki sadece bir tane en yakın kelimeye ait olur. Bu iki nedenden ötürü çok büyük nicemleme hataları oluşur.

Yumuşak nicemleme/eşleme yöntemi, her bir betimleyici vektörü en yakın birden fazla görsel kelimeye atar. Atama ağırlıklarını yerel betimleyici uzayındaki sözcük merkezlerine olan yakınlığına göre ayarlar. Bu yöntem orijinal görüntü öznitelikleri hakkında daha fazla bilgi saklar. Sabit eşlemedeki sorunları görsel kelimelere ağırlıklı atama yaparak azaltır.

KKT yumuşak eşleme uygulayan bir kodlama yöntemidir. KKT'de vektör nicemleme için, tek bir kodsözcüğüne atama (sabit nicemleme) yerine birden fazla kodsözcüğüne yumuşak eşleme yapılmaktadır. Kodsözcüklerine atanan ağırlıklar bir kernel uzaklık fonksiyonu ile hesaplanmaktadır. Betimleyicilerin kodlanması daha detaylı olarak van Gemert vd. (2008)'de şu şekilde verilmiştir ( $\mathbf{x}$  öznitelik vektörü ve  $\mathbf{a}_t$  kodsözcüğü olsun,  $1 \leq t \leq A$ ):

$$[f_{kkt}(\mathbf{x})]_t = \frac{K(\mathbf{x}, \mathbf{a}_t)}{\sum_{j=1}^A K(\mathbf{x}, \mathbf{a}_j)} \quad (3.15)$$

Burada

$$K(\mathbf{x}, \mathbf{a}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{a}\|^2}{2\sigma^2}\right) \quad (3.16)$$

KKT kodlamada daha iyi bir verim alabilmek adına öznitelik vektörlerinin en yakın 5 görsel kelime ile şifrenmesi önerilmiştir. Dolayısıyla  $A = 5$  ve  $\{\mathbf{a}_t\}_{1 \leq t \leq A}$   $\mathbf{x}$  vektörüne en yakın görsel kelimelerin kümesidir. KKT'de  $\sigma$  parametresi ağırlıkların ne kadar homojen dağıtıldığını belirlemektedir;  $\sigma$  arttıkça ağırlıklar birbirine yaklaşmaktadır. Özet olarak, kernelin şekli direkt olarak uzaklık fonksiyonundan gelirken, kernelin boyutu veriye ve görüntü betimleyiciye bağlıdır.

Kod-tablosundaki toplam görsel kelime sayısı  $T$  olsun. KKT kodlama sonucu  $\mathbf{f}_{kkt}(\mathbf{x}) = [[\mathbf{f}_{kkt}(\mathbf{x})]_1 \dots [\mathbf{f}_{kkt}(\mathbf{x})]_T]$  kodvektörü  $T$  uzunluğunda ve  $A = 5$  en yakın görsel kelime dışında kalan katsayıları sıfırdır. Ayrıca  $\|\mathbf{f}_{kkt}(\mathbf{x})\|_1 = 1$ .

KKT betimleyici için son adım, süperpiksele ait kodvektörlerin birleştirilmesidir. Bunun için en çok tercih edilen yöntemler ortalama birleştirme ve maksimum birleştirmedir. Ortalama birleştirme operasyonu, süperpiksele ait bütün yerel betimleyicilerin katsayılarını toplar ve düzgeler:

$$\mathbf{f}_{kkt} = \frac{1}{|SP|} \sum_{i \in SP} \mathbf{f}_{kkt}(\mathbf{x}_i) \quad (3.17)$$

Burada  $|SP|$  süperpiksel boyutunu (yani piksel sayısını) temsil etmektedir. Dolayısıyla ortalama birleştirme ile öznitelik histogramı elde edilmektedir.

### Yerellik-Kısıtlı Doğrusal Kodlama (YDK)

YDK yönteminde de, KKT'de olduğu gibi, öznitelik vektörü yerel komşuluğunda yer alan kodsözcüklerinin doğrusal bir birleşimi olarak ifade edilir. Fakat kodsözcüklerine atanan ağırlıkların tespiti için farklı bir yaklaşım izlenir. KKT'de olduğu gibi  $\{\mathbf{a}_t\}_{1 \leq t \leq A}$   $\mathbf{x}$  öznitelik vektörüne en yakın görsel kelimelerin kümesi,  $A = 5$  ve  $\mathbf{B} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_A]$  sütunları bu en yakın kodsözcüklerinden oluşan matris olsun. Kodsözcüklerine atanan ağırlıklar  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_A]^T$  aşağıdaki eniyileme probleminin çözümü ile belirlenir:

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\mathbf{1}^T \boldsymbol{\alpha} = 1} \|\mathbf{x} - \mathbf{B}\boldsymbol{\alpha}\|^2 + \beta \|\boldsymbol{\alpha}\|^2 \quad (3.18)$$

Burada  $\beta$  küçük bir düzleştirme sabitidir ve  $\mathbf{1}^T \boldsymbol{\alpha} = 1$  kısıtı ile ağırlıkların toplamının 1 olması sağlanır.

Kod-tablosundaki toplam görsel kelime sayısı  $T$  ise, YDK kodlama sonucu elde edilen betimleyici  $\mathbf{f}_{ydk}(\mathbf{x}) = [[\mathbf{f}_{ydk}(\mathbf{x})]_1 \dots [\mathbf{f}_{ydk}(\mathbf{x})]_T]$  kodvektörü  $T$  uzunluğunda olacaktır ve sadece  $A = 5$  en yakın görsel kelimeye karşılık gelen katsayıları sıfırdan farklıdır:  $[\mathbf{f}_{ydk}(\mathbf{x})]_{I_{\mathbf{a}_t}} = \alpha_t$ ,  $1 \leq t \leq A$  ( $I_{\mathbf{a}_t}$ , görsel kelime  $\mathbf{a}_t$ 'nin sözlükteki sıra numarasıdır).

YDK betimleyici için son adım, süperpiksele ait kodvektörlerin birleştirilmesidir. Bunun için YDK'da maksimum birleştirmenin, ortalama birleştirmeye göre daha başarılı olduğu gösterilmiştir. Dolayısıyla,

$$[\mathbf{f}_{ydk}]_t = \max_{i \in SP} [\mathbf{f}_{ydk}(\mathbf{x}_i)]_t, \quad 1 \leq t \leq T \quad (3.19)$$

Son olarak süperpiksel kodvektörü normu 1 olacak şekilde düzgelenir:

$$\mathbf{f}_{ydk}^N = \frac{\mathbf{f}_{ydk}}{\|\mathbf{f}_{ydk}\|} \quad (3.20)$$

### 3.4.2 Benzetimler

Bu çalışmada SuperParsing'in altı (6) farklı SIFT özneliği için KKT ve YDK kodlama kullanılması test edilmiştir. Diğer 14 öznelik değiştirilmemiştir. Orijinal algoritmada bunun dışında iki değişiklik daha yapılmıştır:

- 1) Öznelik komşuluğuna geri erişim kümesinden sabit sayıda en yakın  $R$  adet süperpiksel dahil edilmiştir;
- 2) KKT kodlanmış SIFT betimleyicileri arasındaki uzaklık hesabı için  $L_1$  ve Bhattacharyya metrikleri ( $d_B$ ) test edilmiştir:

$$d_B(\mathbf{f}_i, \mathbf{f}_j) = -\ln \left( \sum_{t=1}^T \sqrt{[f_i]_t \cdot [f_j]_t} \right) \quad (3.21)$$

YDK kodlanmış betimleyiciler için ise  $L_2$  uzaklık metriği kullanılmıştır.

Benzetimlerde orijinal SuperParsing (**SP**) ile dört modifiye versiyonu karşılaştırılmıştır: **SPv1** (sadece 1 no'lu değişiklik), **SP+KKT+L1** (KKT kodlama, 1 ve 2 no'lu değişiklikler,  $L_1$  uzaklık metriği), **SP+KKT+BH** (KKT kodlama, 1 ve 2 no'lu değişiklikler, Bhattacharyya uzaklık metriği), **SP+YDK+L2** (YDK kodlama, 1 no'lu değişiklik,  $L_2$  uzaklık metriği). Ayrıca KKT kodlamada sözlük büyüklüğü  $T$  ve  $\sigma$  parametresi için farklı alternatifler test edilmiştir:

- i) **SP+KKTv1**: SP+KKT+L1,  $T=256$ ,  $\sigma = 75$
- ii) **SP+KKTv2**: SP+KKT+L1,  $T=512$ ,  $\sigma = 100$
- iii) **SP+KKTv3**: SP+KKT+L1,  $T=1024$ ,  $\sigma = 100$
- iv) **SP+KKTv4**: SP+KKT+BH,  $T=256$ ,  $\sigma = 75$
- v) **SP+KKTv5**: SP+KKT+BH,  $T=512$ ,  $\sigma = 100$

YDK kodlama için de  $T$  ve  $\beta$  parametresi şu şekilde seçilmiştir:

- i) **SP+YDKv1**: SP+YDK+L2,  $T=256$ ,  $\beta = 0.015$
- ii) **SP+YDKv2**: SP+YDK+L2,  $T=512$ ,  $\beta = 0.008$
- iii) **SP+YDKv3**: SP+YDK+L2,  $T=1024$ ,  $\beta = 0.006$

Algoritmaların başarımı SIFT Flow veri kümesi üzerinde test edilmiştir. Benzetimlerde piksel-seviyesinde sınıflandırma doğruluğu (yani, doğru sınıflandırılan piksel yüzdesi) yanında sınıf seviyesinde doğruluk ortalamalarının karşılaştırılması yapılmıştır. SuperParsing'de ÇTB parametre değerleri  $K = 400, S = 200$  ve komşuluk boyutu  $R = 15$ .

Tablo 3.4. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Temel	MRA
SP	73.9 / 29.5	75.4 / 27.1
SPv1	76.1 / 27.2	76.6 / 23.3
SP+KKTv1	77.1 / 29.9	77.6 / <b>29.0</b>
SP+KKTv2	<b>77.8 / 30.7</b>	78.0 / <b>29.0</b>
SP+KKTv3	77.4 / 30.6	77.6 / 28.3
SP+KKTv4	77.2 / 30.1	<b>78.1 / 27.3</b>
SP+KKTv5	77.6 / 30.5	77.7 / 28.7
SP+YDKv1	77.7 / 29.4	<b>78.1 / 27.9</b>
SP+YDKv2	75.9 / 28.9	76.6 / 27.2
SP+YDKv3	77.0 / 29.6	77.3 / 27.4

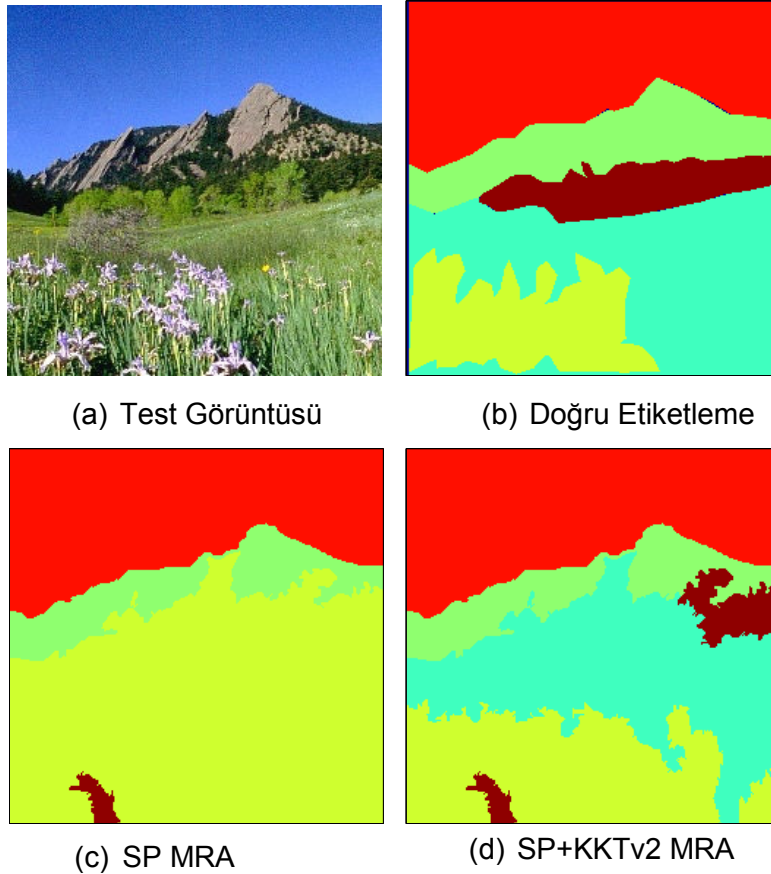
Tablo 3.4'te toplam 9 versiyon ile birlikte SuperParsing algoritmasının piksel etiketleme ve ortalama sınıf doğruluk yüzdeleri verilmiştir. Algoritmaların Temel ve MRA versiyonları için karşılaştırma yapılmıştır. KKT ve YDK kodlama ile SPv1 Temel algoritmaya göre piksel doğruluğunda %1.7'e varan artış gözlenmiştir. SPv1 SuperParsing'e göre piksel doğruluğunda %2.2 daha yüksek iken sınıf ortalamalarında daha düşük sonuç vermiştir. KKT ve YDK kodlama kullanılarak her iki ölçütte de iyileşme sağlanmıştır. Sadece SP+YDKv2 sonuçları beklentinin altındadır. MRA modeli kullanıldığında piksel doğruluğunda artışlar daha kısıtlı kalmıştır. SPv1 MRA ile karşılaştırıldığında piksel doğruluğunda %1.5, sınıf ortalamalarında %5.7'e varan artış sözkonusudur. SP MRA ile karşılaştırıldığında ise piksel doğruluğunda %2.7, sınıf ortalamalarında %1.9'a varan artış elde edilmiştir. Sözlük büyüklüğü ve uzaklık metriği karşılaştırmasında net bir sonuç gözlenmemiştir. KKT'de  $T=512$  ve  $L_1$  uzaklığı (SP+KKTv2) için biraz daha iyi sonuçlar elde edildiği söylenebilir.

Tablo 3.5, 20 öznelik yerine tek bir SIFT betimleyicinin kullanılması durumunda elde edilen sonuçları göstermektedir. Bunun için en başarılılardan olan süperpikselin tüm pikselleri kullanılarak hesaplanan SIFT betimleyici tercih edilmiştir. Görüldüğü gibi KKT kodlama SIFT betimleyicinin etiketleme doğruluğunda %6.4'e varan artış sağlamaktadır. Sınıf ortalamalarında artış %9.1'e kadar çıkmaktadır. Bu sonuçlarda Bhattacharyya metriğinin  $L_1$ 'den piksel doğruluğu ölçütünde daha başarılı olduğu görülmektedir. MRA ile Temel

algoritmaya göre yaklaşık %10'luk artış elde edilmekte ve tek bir betimleyici ile %70'in üzerinde piksel etiketleme doğruluğu sağlanmaktadır.

Tablo 3.5. SIFT betimleyici için doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Temel	MRA
SP	46.7 / 23.0	60.9 / 23.3
SPv1	55.2 / 24.7	66.2 / 24.9
SP+KKTv1	59.4 / 33.8	70.0 / <b>33.7</b>
SP+KKTv2	59.8 / 31.9	69.1 / 29.0
SP+KKTv3	59.8 / 33.8	69.5 / 31.7
SP+KKTv4	<b>61.6</b> / 33.1	<b>71.8</b> / 30.8
SP+KKTv5	<b>61.6</b> / 33.1	71.0 / 32.3
SP+YDKv1	60.7 / 33.1	69.5 / 32.2
SP+YDKv2	59.8 / 33.0	68.7 / 31.2
SP+YDKv3	61.0 / <b>34.1</b>	68.9 / 32.5



Şekil 3.4. Görsel karşılaştırma (kırmızı: gökyüzü, yeşil: dağ, k.rengi: ağaç, mavi: yeşil alan, sarı: bitki)

Şekil 3.4'te SP MRA ve SP+KKTv2 MRA için etiketleme doğruluğu görsel olarak karşılaştırılmıştır. SuperParsing oldukça zor olan *yeşil alan-bitki* sınıf ayrımını başaramazken KKT kodlama ile bu mümkün olmuş, hatta *bitki* sınıfı doğru etiketlemeden daha gerçekçi şekilde ayrıştırılmıştır. Ayrıca dağın eteklerinde zor görünen *ağaç* sınıfını da bir miktar tespit edebilmiştir. Sonuçlar görüntüden görüntüye değişiklik göstermekle birlikte genel gözlem KKT ve YDK kodlama ile daha tutarlı sonuçlar üretildiğidir.

Sonuç olarak, benzetim sonuçları KKT ve YDK ile öznitelik uzayında farklı sınıfların koşullu dağılımlarının birbirinden daha iyi ayrıldığı ve böylece daha doğru sınıf olabilirlik oranları kestirildiğini göstermiştir.

### 3.5 CNN Özniteliklerinin SuperParsing'de Kullanılması

Derin evrimsel sinir ağları (CNN) pekçok sınıflandırma probleminde olduğu gibi görüntü sınıflandırma ve sahne etiketleme alanlarında da çok ciddi başarı göstermiştir. CNN mimarisinin en önemli özelliklerinden birisi görüntüyle ilgili herhangi bir ön-işleme ve betimleyici tanımlamaya gerek olmadan sınıflandırıcı eğitiminin yapıyor olmasıdır. Dolayısıyla CNN mimarisi sınıflandırma için gerekli görüntü özniteliklerini de eğitim sırasında öğrenmektedir. Çok katmanlı bu mimarilerin farklı katmanlarından elde edilen vektörel değerler görüntü ile ilgili çeşitli özniteliklere denk gelmektedir. Çok geniş görüntü veri setlerinde eğitilen derin mimarilerin katmanlarından çıkarılan bu öznitelikler çeşitli sınıflandırma problemlerinde kullanılmış ve SIFT gibi “el yapımı” özniteliklerden daha başarılı sonuçlar vermiştir (Budnik vd., 2015). Ayrıca öğrenilmiş özniteliklerle “el yapımı” özniteliklerin birlikte kullanılmasının da sınıflandırma doğruluğunu artırdığı gösterilmiştir (Budnik vd., 2015; Park vd., 2016).

Bu bölümde SuperParsing için CNN özniteliklerinin kullanılmasıyla ilgili yapılan çalışmalardan bahsedilmiştir. Nguyen vd. (2016)'dakine benzer şekilde, CNN öznitelikleri hem geri erişim kümesi belirlerken global eşleştirme için ve hem de yerel süperpiksel eşleştirme için kullanılmıştır. Kullandığımız öznitelikler, VGG-F (Chatfield vd., 2014) ve AlexNet (Krizhevsky vd.,2012) ağlarının en son sınıflandırma katmanından önceki katmandan elde edilmiştir. Bu derin ağlar ILSVRC ImageNet veri kümesinde (Deng vd., 2009) eğitilmiştir.

Geri erişim kümesi tespitinde, test görüntüsü için CNN ağ katmanından 4096 boyutlu bir öznitelik vektörü elde edilir. Bu vektör düzgelendikten sonra tüm eğitim görüntülerinin CNN vektörlerine olan  $L_2$  uzaklığı hesaplanır. Kullanılan diğer global betimleyicilerle birlikte Bölüm 2.1'de anlatıldığı gibi bir sıralama ile geri erişim kümesi belirlenir.

Süperpikseller için, 20 adet özniteliğin yanında, VGG-F ve AlexNet öznitelikleri için de sınıf-koşullu olabilirlik oranları Bölüm 2.1'deki gibi hesaplanmıştır. Yerel seviyede CNN

özniteliği, süperpiksel bölgesi hariç tüm görüntünün sıfıra atanması ve ağ çıktısının bu “süperpiksel görüntüsü” için hesaplanması ile bulunur.

Öğrenilmiş CNN özniteliklerinin diğer özniteliklerle kaynaştırılması SuperParsing’in ayırıştırma doğruluğunu artırmaktadır. Fakat Bölüm 3.3’te de tartışıldığı gibi bazı özniteliklerin SuperParsing performansına katkısı düşüktür ve CNN öznitelikleri ile kullanılmalarının faydası olmayabilir. Bu nedenle, hem global betimleyiciler hem de süperpiksel betimleyiciler için, farklı öznitelik kombinasyonları denenmiş ve en iyi performans veren öznitelik altkütmesi bulunmaya çalışılmıştır. Ayrıca daha başarılı olan (KCB/LLC kodlanmış) SIFT ve CNN özniteliklerinin log-olabilirlik oranlarına daha yüksek ağırlık,  $w_k$ , atanarak etiketleme başarımının nasıl değiştiği analiz edilmiştir:

$$L(s_i, c) = \sum_k w_k \log \frac{P(f_i^k | c)}{P(f_i^k | \bar{c})} \quad (3.22)$$

Tablo 3.6 ve 3.7’de SIFT Flow veri kümesinde farklı öznitelik kombinasyonları ve ağırlıklandırmaları ile elde edilen sonuçlar iki bölütleme alternatifi için sunulmuştur (( $K=400$ ,  $S=200$ ,  $R=15$ ) ve ( $K=200$ ,  $S=100$ ,  $R=30$ )). Yöntemlerle ilgili kullanılan kısaltmalar şu şekilde tanımlanmıştır:

- GB (Global betimleyici): VGG-F, AlexNet, uzamsal piramit (UP), Gist, renk histogramı (RH)
- SB (Süperpiksel betimleyici):
  - Tümü: 22 öznitelik (VGG-F, AlexNet ve standart 20 öznitelik)
  - Kısıtlı: 18 öznitelik (kullanılmayanlar: renk standard sapması, 2 tekton özniteliği, kuşatan kutu (bounding box) boyutu)
  - LLC/KCB: 6 adet SIFT özniteliği için LLC/KCB kodlama
  - Ağırlıklı: VGG-F, AlexNet için  $w_k = 2$ ; ortalama renk ve 2 SIFT özniteliği için  $w_k = 3$ ; diğerleri için  $w_k = 1$ .

Tablolardan ortaya çıkan sonuca göre, süperpikseller için CNN özniteliklerinin kullanılması piksel doğruluğunu her iki alternatif için de %0.7 artırmıştır. Geri erişim kümesinin belirlenmesinde VGG-F ve AlexNet özniteliklerinin global betimleyici olarak kullanılması ise hem piksel doğruluğunda hem de ortalama sınıf doğruluklarında kayda değer artışlar sağlamıştır. Bu sonuç, test görüntüsüne daha benzer eğitim seti görüntülerinden gelen süperpiksellerle eşleştirme yapılmasının daha doğru log-olabilirlik oranı hesabı için önemini ortaya koymaktadır. Geri erişim kümesinin iyileştirilmesi, test görüntüsünde varolan nadir sınıflarla etiketlenmiş benzer eğitim süperpikselleri bulma olasılığını artırmakta; bu da

nadir sınıfların etiketlenme doğruluğunu ve dolayısıyla ortalama sınıf doğruluğunu iyileştirmektedir. Global betimleyici olarak VGG-F veya AlexNet ile beraber uzamsal piramidin de kullanılması başarılı sonuçlar vermiştir. Dolayısıyla CNN öznitelikleri ile “el yapımı” özniteliklerin birlikte kullanılmasının performansa olumlu katkısı vardır. Süperpikseller için bazı düşük performanslı özniteliklerin çıkarılmasının performansa büyük bir katkısı olmamıştır. Log-olabilirlik oranlarının ağırlıklandırılması ise piksel doğruluğunu bir miktar artırırken sınıf ortalamasını düşürebilmektedir. Sonuç olarak, yapılan tüm denemeler sonucunda, piksel doğruluğunda %3.6, sınıf ortalamasında ise %8.6'ya varan iyileşmeler gözlenmiştir.

Tablo 3.6. SIFT Flow veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları ( $K=400$ ,  $S=200$ ,  $R=15$ )).

Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
SP+KKTv2	78.0	29.0
GB: UP, Gist, RH SB: Tümü, KCB	78.7	29.8
GB: UP, Gist, AlexNet SB: Tümü, Ağırlıklı, KCB	81.2	32.9
GB: UP, RH, VGG-F SB: Kısıtlı, Ağırlıklı, KCB	81.5	30.4
GB: UP, AlexNet SB: Kısıtlı, KCB	81.6	35.7
GB: UP, VGG-F SB: Kısıtlı, KCB	81.4	37.6
GB: UP, AlexNet SB: Kısıtlı, Ağırlıklı, KCB	81.6	32.3

Tablo 3.7. SIFT Flow veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları ( $K=200$ ,  $S=100$ ,  $R=30$ )).

Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
SP+YDKv1	78.2	30.8
GB: UP, Gist, RH SB: Tümü, LLC	78.9	31.1
GB: UP, Gist, AlexNet SB: Tümü, LLC	80.3	35.2
GB: UP, VGG-F SB: Kısıtlı, LLC	80.5	36.2
GB: UP, RH, AlexNet SB: Kısıtlı, Ağırlıklı, LLC	80.9	35.6

Benzer bir çalışma, 19-sınıflı LabelMe veri kümesinde (Jain vd. (2010)) de yürütülmüştür. Bu veri kümesinde 250 eğitim, 100 test görüntüsü vardır. Açık hava



görüntülerinden oluşan bu kümede *ağaç, taş, uçak* gibi 19 farklı sınıf yer almaktadır. Eğitim kümesi küçük olduğundan geri erişim kümesi için 50 ve 100 adet görüntü kullanılması test edilmiştir (EK=50 veya 100).

Tablo 3.8. LabelMe veri kümesinde CNN öznitelikleri ile SuperParsing benzetim sonuçları (( $K=200$ ,  $S=100$ ,  $R=30$ )).

Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
SP+KKTv2 (EK:100)	81.4	54.1
GB: UP, RH, VGG-F SB: Kısıtlı, Ağırlıklı, KCB EK: 100	85.4	62.3
GB: UP, VGG-F SB: Kısıtlı, KCB EK: 100	85.8	62.4
GB: UP, AlexNet SB: Kısıtlı, Ağırlıklı, KCB EK: 50	86.7	67.0

Tablo 3.8'de LabelMe veri kümesinde kısıtlı öznitelik kümesi ve KCB kodlama için farklı benzetim sonuçları sunulmuştur. Geri erişim kümesinin belirlenmesinde uzamsal piramit ile birlikte VGG-F özniteliği kullanılması ve süperpikseller için VGG-F/AlexNet özniteliklerinin eklenmesi piksel doğruluğunda %4.4, sınıf ortalamasında %8.3 artış sağlamıştır. Geri erişim kümesinin 100 yerine 50 eğitim görüntüsüne kısıtlanması ve VGG-F yerine Alex özniteliğinin kullanılması da piksel ve sınıf doğruluklarında, sırasıyla, %0.9 ve %4.6 iyileşmeye neden olmuştur. Bu sonuçlar sadece 250 adet eğitim görüntüsü içeren küçük bir veri seti olan LabelMe'de test görüntüsü ile iyi eşleşen sınırlı bir altküme belirlenmesinin SuperParsing algoritmasının ayrıştırma başarımı açısından önemini ortaya koymaktadır.

## 4. ÇOK-HİPOTEZLİ BAĞLAMSAL MODELLEME

Projenin ikinci yılında SuperParsing için daha gelişmiş bağlamsal modeller geliştirerek ayrıştırma doğruluğunun iyileştirilmesi hedeflenmiştir. Bölüm 3'te detaylandırılan çalışmalar, süperpiksel bölütleri için uyarlanabilir öznitelik seçimi, olabilirlik oran hesabı ve ağırlıklandırmasının etiketleme başarımı üzerindeki etkilerini ortaya koymuştu. Bu bölümde ise süperpiksel komşuluk bilgisinin farklı bölütleme alternatiflerinin kaynaştırılmasında etkin olarak kullanılmasını sağlayan esnek bir yaklaşım sunulmuştur.

Çok-hipotezli SuperParsing için geliştirilmiş bir bağlamsal model olarak tasarlanmış bu yaklaşım "çok-hipotezli MRA" olarak adlandırılmıştır. Bu MRA modeli ile daha uyarlanabilir komşuluk tanımları elde edilmiş ve etiketleme başarımında artış sağlanmıştır. Yapılan çalışmaların bir kısmı ile bir konferans bildirisi yayınlanmıştır (Ates ve Sunetci, 2017b). Daha sonra, yapılan çalışmalar CNN özniteliklerinin kullanılması ve FCN bölütlemenin de dahil edilmesi ile genişletilmiş; bu çalışmalardan ortaya çıkan dergi makalesi (Ates ve Sunetci, 2017c) hakem değerlendirmesine gönderilmiştir.

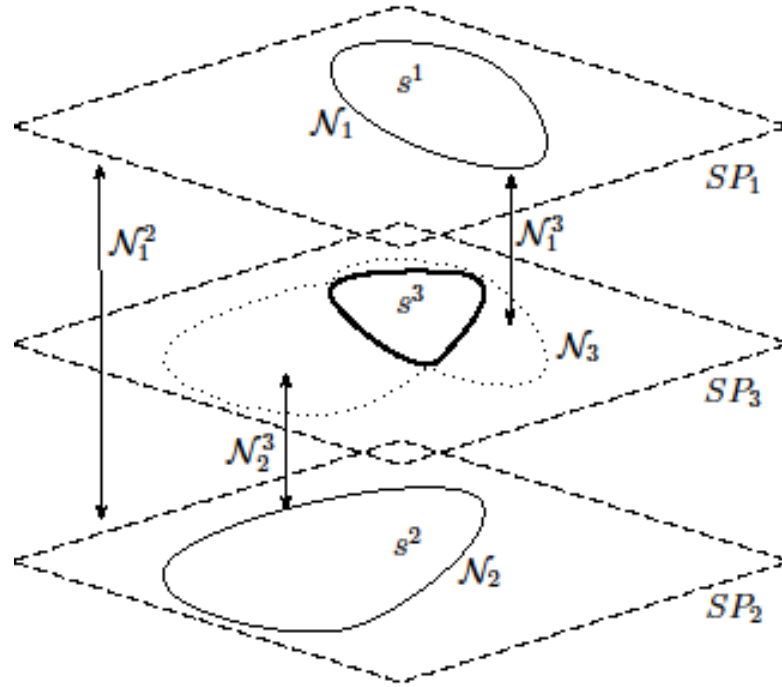
Aşağıda çok-hipotezli MRA modeli ve yürütülen benzetimler detaylı olarak açıklanmıştır.

### 4.1 Alternatif Bölütlemeler için MRA Bağlamsal Modeli

Süperpikseller nesne sınırlarıyla genelde uyumlu olmakla birlikte sabit bir bölütleme kullanılması, görüntüdeki bağlamsal bilginin temsili için oldukça kısıtlayıcıdır. Ayrıştırma performansı, süperpiksellerin boyut, şekil, sınır ve içeriğinin görüntüde varolan nesne sınıflarını ne kadar temsil edebildiğine oldukça bağlıdır. Farabet vd. (2013) çoklu-ölçekli bölütler kullanarak hiyerarşik öznitelikler öğrenir ve birçok alternatiften eniyi bir bölüte ulaşmaya çalışır. Fakat sonuç olarak bu yöntem eniyi olduğunu iddia ettiği tek bir bölütlemeye bağlı kalmakta ve alternatif gösterimlerin ortak eniyilenmesine çalışmamaktadır. FCN derin mimarisi (Shelhamer vd., 2016) farklı çözünürlükteki bağlamsal bilginin kaynaştırılması için kaba katman öznitelikleri ince, alt-katman özniteliklerle birleştirmektedir. Ne var ki, ağ katmanları dikkörtgensev evrişim ve düzenli örnekleme ile türetildiği için, bu katmanlardaki öznitelikler sahnedeki nesnelerin geometrisi ile uyuşmamaktadır.

Proje öncesi çalışmamızda (Ak ve Ateş, 2015) çok-hipotezli SuperParsing ile birden fazla alternatif bölütleme kullanılmasının ayrıştırma doğruluğunu artırdığını göstermiştik. Nesne sınıflarının çok farklı boyutları (ör. *gökyüzü-pencere*) ve farklı dokuları (ör. *ağaç-bina*) olması sebebiyle alternatif süperpiksel gösterimleri fayda sağlamaktadır.

Bu bölümde açıklanan çalışmada ise çok-hipotezli SuperParsing için özgün bir MRA modeli tanımladık. Bu MRA modeli için tanımlanan yerel bağlam sadece aynı bölütlemedeki bitişik süperpiksellerin uzamsal komşuluğunu dikkate almamakla kalmayıp farklı bölütlemeler arasında kesişen süperpiksellerin komşuluğunu da içermektedir. Bu modelin dayandığı temel iddia farklı bölütlemelerin görüntüdeki nesne sınıfları ile ilgili tümler (complementary) bilgi sağladığıdır. Önerilen yaklaşım alternatif bölütlemelerde varolan bilginin kaynaştırılması için tümleşik bir model oluşturmaktadır. Başka deyişle önerilen MRA modeli, daha tutarlı etiketleme için, hem uzamsal hem de bölütlemeler-arası süperpiksel komşuluklarında bağlamsal kısıtlar kodlanması için kullanılmıştır. Sonuç olarak, tek bir bölütlemenin sabit bağlamına göre, komşuluk bağlamının daha esnek ve geliştirilmiş bir gösterimi elde edilmiştir.



Şekil 4.1. Çok-hipotezli MRA modeli

Aşağıda, MRA modeli iki alternatif bölütleme için tanımlanmıştır; fakat herhangi sayıda alternatif için kolayca geliştirilebilir. Yukarıda belirttiğimiz gibi, bu MRA modelinde hem bölütleme-içi hem de bölütlemeler-arası bağlamsal bilgi tanımlanmaktadır. Bölütleme-içi komşuluk daha önceden olduğu gibi ilgili bölütlemenin bitişik süperpiksellerini içerir. Bölütlemeler-arası komşuluk ise farklı bölütlemelerden kesişen süperpikselleri içerir. Süperpiksel kümesi  $SP_m = \{s_i^m\}$  ( $m = 1,2$ ) olarak tanımlansın. İki alternatifin süperpiksellerinin kesişimine bağlı olarak üçüncü bir bölütleme  $SP_3 = \{s_k^3\}$  tanımlıyoruz (bkz. Şekil 4.1):

$$s_k^3 = s_i^1 \cap s_j^2 \neq \emptyset, \quad \forall s_i^1 \in SP_1, s_j^2 \in SP_2 \quad (4.1)$$

$N_m$  bir bölütlemedeki bitişik süperpiksel çiftlerini içeren bağlamsal komşuluğu temsil etsin. Bu bölütleme-içi komşuluklara ek olarak, bölütleme-arası bağlamsal komşulukları  $N_n^m$  ( $n, m = 1, 2, 3$ ) şu şekilde tanımlıyoruz:

$$(s_i^n, s_j^m) \in N_n^m \quad \text{eğer} \quad s_i^n \cap s_j^m \neq \emptyset \quad \text{ise} \quad (4.2)$$

Her  $s_i^m$  süperpikseline bir  $c_i^m$  sınıf etiketi atanarak görüntü ayrıştırılmaktadır. Dolayısıyla her bölütleme alternatif bir ayrıştırma sonucu,  $\mathbf{c}^m = \{c_i^m\}$  üretir. Görüntü etiketlemeyi tüm süperpiksel etiket kümesi,  $\mathbf{c} = \{\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3\}$ , üzerinde tanımlı bir MRA enerji enküçültme problemi olarak tanımlıyoruz:

$$J(\mathbf{c}) = \sum_{m=1}^3 \left( \sum_{s_i^m \in SP_m} D(s_i^m, c_i^m) + \lambda_m \sum_{(s_i^m, s_j^m) \in N_m} E(c_i^m, c_j^m) \right) + \sum_{(n,m) \in IC} \lambda_n^m \sum_{(s_i^n, s_j^m) \in N_n^m} E(c_i^n, c_j^m) \quad (4.3)$$

Burada  $D$  ve  $E$ , sırasıyla veri ve düzlük masraflarına karşılık gelmektedir.  $IC$  bölütleme-arası komşuluk kümesidir:  $IC = \{(1,2), (1,3), (2,3)\}$ .  $\lambda_m$  ve  $\lambda_n^m$  ise ilgili komşuluklar için düzlük sabitleridir.

Yukarıdaki MRA formülasyonunda  $D(s_i, c)$  veri masrafı süperpiksel  $s_i$ 'nin  $c$  sınıfına ait olma ihtimalini temsil etmektedir.  $E(c_i, c_j)$  düzlük masrafı ise iki komşu süperpikselin farklı sınıf etiketlerine ( $i \neq j$ ) sahip olma olasılığını ölçmektedir. Herhangi bir ayrıştırma algoritması, eğer veri masrafı olarak kullanılabilir sınıf-koşullu olabilirlik skorları üretiyorsa, bu MRA formülasyonu içerisinde kullanılabilir. Bölüm 4.2'de  $SP_1$  ve  $SP_2$  alternatif bölütleri SuperParsing'in farklı parametre kümeleri için çalıştırılması ile edilmiştir. Bölüm 4.2.1'de ise ikinci alternatif için FCN bölütleme kullanılması durumu incelenmiştir.  $SP_1$  ve  $SP_2$ 'nin düzlük masrafları için ise Denklem (2.5)'teki model kullanılmıştır.

$SP_3$  diğer iki bölütlemeden elde edildiği için, veri ve düzlük masrafları da  $SP_1$  ve  $SP_2$ 'deki karşılık gelen masrafların birer fonksiyonu olarak tanımlanmıştır. İki bölütlemeden gelen tümleyici bilgiyi kaynaştırabilmek için,  $SP_1$  ve  $SP_2$  bölütlerindeki sınıf kümelerini birbirinden ayırıyoruz ve  $SP_3$  bölütü için bu iki kümenin birleşiminden seçim yapıyoruz. Diğer bir ifadeyle,  $SP_1$  ve  $SP_2$ 'deki  $c$  anlambilimsel sınıfına karşılık gelen  $SP_3$ 'te iki farklı etiket

tanımlıdır:  $c^{(1)}$  ve  $c^{(2)}$  (Burada  $c_i^m$  etiketi  $s_i^m$  süperpikseline atanan sınıfı temsil ederken  $c^{(m)}$  ise  $SP_m$ 'in herhangi bir sınıfını işaret etmektedir). Buna göre  $s_k^3 \in SP_3$  için veri masrafını şöyle ifade ediyoruz ( $m = 1,2$ ):

$$D(s_k^3, c^{(m)}) = f_m(D(s_i^1, c), D(s_j^2, c)) \quad (4.4)$$

Burada  $(s_i^1, s_k^3) \in N_1^3$  ve  $(s_j^2, s_k^3) \in N_2^3$ . Benzer şekilde, bölütleme-arası komşuluk  $N_n^m$  için düzlük maliyeti  $N_1$  ve  $N_2$ 'deki maliyete bağlı hesaplanır:

$$E(c_i^{(1)}, c_j^{(2)}) = g(E(c_i^{(1)}, c_j^{(1)}), E(c_i^{(2)}, c_j^{(2)})) \quad (4.5)$$

$f_m$  ve  $g$  fonksiyonları detaylı olarak aşağıdaki bölümde açıklanmıştır.

$\lambda_m$  ve  $\lambda_n^m$  düzlük sabitleri Denklem (4.3)'teki MRA modelinde farklı komşuluklardaki bağlamsal bağımlılıkların seviyesini kontrol etmektedir. Bu değerler eğitim kümesinden birini-dışarıda-bırak (leave-one-out) stratejisi ile tespit edilmiştir: her bir eğitim görüntüsü eğitim kümesinden çıkarılır ve farklı parametre değerleri için önerilen yöntem ile etiketlenir. Daha sonra eğitim kümesindeki ortalama ayrıştırma doğruluğunu eniyileyen parametre kümesi belirlenir.

MRA enerji fonksiyonu, Boykov ve Kolmogorov (2004)'deki  $\alpha$ -genleşme metodu ile minimize edilir. Sonuçta üç alternatif etiketleme elde edilir:  $c = \{c^1, c^2, c^3\}$ . Görüntünün final etiketlemesi olarak  $SP_3$  bölütlemesinin etiket kümesi  $c^3$  seçilmiştir.

## 4.2 Çok-hipotezli MRA için Benzetimler

MRA modelinin düzlük sabitleri şu kümeden seçilmiştir:  $\{\lambda | l \in \{0,1,2\}; 5 \leq \lambda \leq 25, \lambda \in \mathbb{Z}\}$ . Ayrıca,  $g(x, y) = 0.5x + 0.5y$ .  $SP_3$ 'ün veri masrafları için ise üç ayrı alternatif test edilmiştir:

- **DC1:** ( $m = 1,2$ )

$$D(s_k^3, c^{(m)}) = \beta_1 \frac{w_k}{w_i^1} D(s_i^1, c) + \beta_2 \frac{w_k}{w_j^2} D(s_j^2, c) \quad (4.6)$$

- **DC2:**

$$\begin{aligned} D(s_k^3, c^{(1)}) &= \frac{w_k}{w_i^1} D(s_i^1, c) \\ D(s_k^3, c^{(2)}) &= \frac{w_k}{w_j^2} D(s_j^2, c) \end{aligned} \quad (4.7)$$

- **DC3:**  $s_k^3$  için veri masraf vektörleri birleştirilerek bir öznitelik vektörü elde edilir:  $\left[ \frac{w_k}{w_i^1} D(s_i^1, \mathbf{c})^T, \frac{w_k}{w_j^2} D(s_j^2, \mathbf{c})^T \right]^T$ . Bu öznitelik vektörü kullanılarak Artırılmış Karar Ağacı (Boosted Decision Tree (BDT)) (Hoiem vd., 2007) sınıflandırıcıları eğitilir. Bu sınıflandırıcının verdiği olabilirlik oran skorları,  $L_{BDT}(s_k^3, \mathbf{c})$ , veri masrafını tanımlamak için kullanılır ( $m = 1, 2$ ):

$$D(s_k^3, c^{(m)}) = 5w_k \sigma \left( L_{BDT}(s_k^3, c) \right) \quad (4.8)$$

Bu modellerde  $w_k = 0.5 |s_k^3| \left( \frac{1}{|s_i^1|} + \frac{1}{|s_j^2|} \right)$  ve  $\beta_1 + \beta_2 = 1$ .

Birinci model  $SP_3$  süperpikselleri için  $SP_1$  ve  $SP_2$ 'de karşılık gelen iki masrafın ortalamasını kullanmakta ve  $c^{(1)}$  ve  $c^{(2)}$  etiketlerini ayırtmamaktadır. İkinci model ise aynı  $c$  sınıfına karşılık gelen  $c^{(1)}$  ve  $c^{(2)}$  etiketlerine iki farklı masraf tanımlamaktadır. Dolayısıyla ikinci model, bağlamsal bilgiye bağlı olarak, MRA modeline iki hipotez arasında seçim yapma imkanı sağlamaktadır.  $w_k$  süperpiksel ağırlığı  $s_k^3$ 'ün  $s_i^1$  ve  $s_j^2$ 'ye göreceli büyüklüğüyle doğru orantılıdır: eğer iki süperpikselin kesişimi küçük ise, o zaman  $s_k^3$ 'ün veri masrafı etiketleme için güvenilir değildir ve bu yüzden daha küçük ağırlık atanır.  $\beta_m$  veri ağırlıkları iki hipotezin bölütleme sonuçlarına olan güveni yansıtmaktadır: daha başarılı algoritmaya daha yüksek ağırlık atanması gerekir.

$SP_1$  ve  $SP_2$  için veri ve düzlük masraflarının belirlenmesinde öncelikle SuperParsing'in farklı parametre değerleriyle koşturulmasıyla elde edilen alternatif gösterimler kullanılmıştır. Bu bölümde bu ilk sonuçlar tartışıldıktan sonra Bölüm 4.2.1'de MRA bağlamsal modeli içerisinde derin CNN özniteliklerinin ve FCN bölütlemenin bir alternatif yöntem olarak kullanılması açıklanmakta ve tartışılmaktadır.

Benzetimlerde test edilen farklı Superparsing alternatifleri şöyle sıralanabilir:

- **SP<sub>1</sub>:**  $(K, S, R) = (200, 100, 30)$ . SIFT öznitelikleri için KCB ya da LLC kodlama.
- **SP<sub>2</sub>:**  $(K, S, R) = (400, 200, 15)$ . SIFT öznitelikleri için KCB ya da LLC kodlama.

Tablo 4.1'de SIFT Flow veri kümesi için MRA modelinde SuperParsing yönteminin kullanılmasıyla elde edilen sonuçlar sunulmuştur. Test edilen veri modelleri:

- **DC1:**  $SP_1$  ve  $SP_2$  için KCB kodlama.  $\beta_1 = \beta_2 = 0.5$
- **DC2:**  $SP_1$  ve  $SP_2$  için KCB kodlama
- **DC3:**  $SP_1$  ve  $SP_2$  için LLC kodlama;  $SP_3$  için KCB kodlama

Tabloda ayrıca karşılaştırma amacıyla, bölütleme-arası komşuluklar olmadan, SuperParsing'in  $SP_1$ (KCB kodlama) ve  $SP_2$ (LLC kodlama) sonuçlarına da yer verilmiştir.

Sonuçları SuperParsing'in dört modifikasyonu ile karşılaştırdık: örnek sezici ile SuperParsing (Tighe, 2015), uyarlanabilir parametrik-olmayan ayrıştırma (Nguyen, 2015), George (2015) ve Kernel+Çok-hipotezli SuperParsing (Ates vd., 2016). (Nguyen, 2015)'de süperpiksel eşleştirmesi ile yerellik-farkındalıklı bir geri erişim kümesi tespit edilip en yakın komşu sayısı  $R$  geri erişim kümesinden uyarlanabilir şekilde belirlenmektedir.

Tablo 4.1. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
DC1	78.5	30.0
DC2	79.5	32.1
DC3	80.6	31.8
$SP_1$	77.8	29.1
$SP_2$	78.1	27.9
SuperParsing	76.2	29.1
Tighe vd. (2015)	78.6	39.2
Nguyen vd. (2015)	78.9	34.0
George (2015)	81.7	50.1
Kernel+Çok-hipotezli	77.8	29.4

DC1 modeli,  $SP_1$  ve  $SP_2$ 'ye göre bölütleme-arası komşulukları kullanarak piksel doğruluğunda %0.5 civarı bir iyileşme sağlamaktadır. DC2, DC1'e göre %1.0 daha yüksek doğruluk sağlayarak iki hipotezdeki tümleyici bilginin kaynaştırılmasında veri masraflarının ayrı tutulmasının katkısını göstermektedir. Başka deyişle, DC2 farklı veri masrafları ve bağlamsal kısıtları olan alternatifler arasında seçim yaparken daha esnek bir gösterim imkanı sağlamaktadır. DC3, DC2'nin piksel doğruluğunu yaklaşık %1 artırmaktadır. Benzetimler bu artışın %0.5'lik kısmının BDT sınıflandırıcı kullanımından, geri kalan %0.5'lik kısmının ise süperpiksel SIFT öznitelikleri için hem LLC hem de KCB kodlama kullanılmasından kaynaklandığını göstermektedir. Burada  $SP_1$  ve  $SP_2$  için LLC kodlama kullanılırken  $SP_3$  için BDT sınıflandırıcıda KCB kodlama ile elde edilen veri masrafları kullanılmaktadır. Her iki kodlama yöntemi tüm veri masraflarının tanımlanmasında birlikte kullanıldığında ise tek bir kodlama yöntemi kullanmaya göre bir kazanç elde edilmemektedir. Bu gözlem geliştirdiğimiz MRA modelinin alternatif kodlama yöntemlerinden gelen tümleyici bilgiyi kaynaştırmayı başardığını ortaya koymaktadır. Önerilen yöntem ortama sınıf doğruluğunda da %3'e varan artış sağlamaktadır.  $c^3$  etiketleri  $SP_1$  ve  $SP_2$  bölütlerinin kesişimi ile oluşan daha ince bir ölçekteki bölütler için ayrıştırma sonucu sunmaktadır. Dolayısıyla önerdiğimiz MRA modeli etiketleme sonucunun fazla düzleşmesini engelleyerek veri kümesindeki nadir sınıflar için doğruluğun artmasına imkan vermektedir.

Sonuç olarak DC3 SuperParsing'e göre piksel doğruluğunda %4.4, ortalama sınıf doğruluğunda ise %2.7 artış sağlamıştır. George (2015) ve Tighe vd. (2015) çalışmalarıyla karşılaştırıldığında, MRA modeli alternatif gösterimlerden gelen olabilirlik skorlarının kaynaştırılması için basit yaklaşım önermektedir. George (2015) Fisher Vektör betimleyicilerin de olduğu yüksek boyutlu öznitelik vektörleri üzerinde eğitilmiş BDT sınıflandırıcıların kaynaştırılmasını önermektedir. Tighe vd. (2015)'de ise SuperParsing olabilirlik oranları veri kümesindeki tüm sınıflar için eğitilen örnek-bazlı SVM sezicilerin sezim skorları ile birleştirilmektedir. MRA modeline göre çok daha karmaşık olan bu çalışmalara yakın sonuçlar elde edilebilmiştir.

Sahne etiketlemede en iyi sonuçlar, öznitelik öğrenme ve olabilirlik kestirimi için derin evrimsel ağlar kullanan yöntemlerle elde edilmektedir. Bölüm 4.2.1'de MRA modelinin derin ağlara uygulanarak nasıl iyileştirilebileceği üzerinde durulacaktır. Bu bölümde ise karmaşık sınıflandırıcı eğitime gerek olmadan, düşük boyutlu öznitelik vektörleri ve veri-güdümlü SuperParsing kullanılarak çok-hipotezli yaklaşımın potansiyeli irdelenmiştir.

Önerilen modellerin ortalama sınıf-seviyesi doğrulukları, Tighe vd. (2015), George (2015) ve literatürde dengeli eğitim seti kullanarak nadir sınıfların doğruluğunu iyileştiren diğer yaklaşımlardan düşüktür. Bu sorunun giderilmesi için dikkatli seçilmiş dengeli bir eğitim seti ile elde edilecek tümleyici bilginin MRA modelinde kullanılmasının faydalı olacağını düşünmekteyiz.

Tablo 4.2'de LabelMe (Jain vd., 2010) veri kümesi için MRA modelinde SuperParsing yönteminin kullanılmasıyla elde edilen sonuçlar sunulmuştur. Karşılaştırma için literatürden Nguyen vd. (2015) ve Myeong ve Lee (2013) sonuçlarına yer verilmiştir. Kullanılan alternatif gösterimler şunlardır:

- **DC2:**  $SP_1$  ve  $SP_2$  için LLC kodlama
- **DC3:**  $SP_1$  ve  $SP_2$  için KCB kodlama;  $SP_3$  için LLC kodlama
- **$SP_1$ :** LLC kodlama.
- **$SP_2$ :** LLC kodlama.

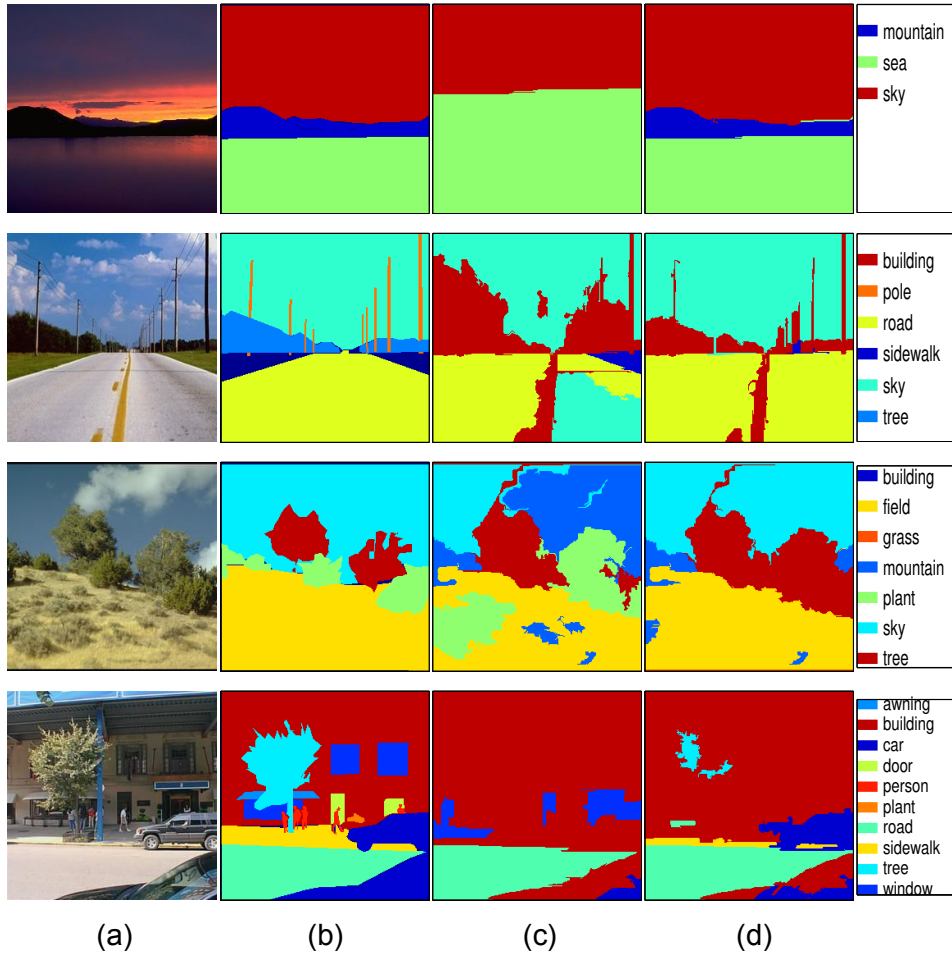
DC3 bu veri setinde de en yüksek sonuçları üretmiştir. Sonuçlarda hem piksel doğruluğu hem de sınıf ortalamalarında iyileşme görülmektedir. Bu veri kümesinde  $SP_2$  sonucu  $SP_1$ 'e göre oldukça düşüktür. Bu da  $c_3$  etiketleme doğruluğunu olumsuz etkilemektedir.  $SP_1$  ve  $SP_2$  için süperpiksel bölütleme parametre değerlerinin daha dikkatli seçilmesi DC3 performansını daha da artırabilir.



Tablo 4.2. LabelMe veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
DC2	83.0	52.9
DC3	83.8	55.6
$SP_1$	81.9	54.7
$SP_2$	79.5	49.9
Myeong ve Lee (2013)	81.8	54.4
Nguyen vd. (2015)	82.7	55.1

Şekil 4.2’de SuperParsing ve DC3 ayrıştırma sonuçları SIFT Flow kümesinden bazı test görüntüleri için karşılaştırılmıştır. DC3 etiketleri genelde SuperParsing’e göre daha tutarlı ve doğrudur. En üstteki şekilde, gözle bile farkedilmesi zor olduğu halde, *dağ* doğru etiketlenmiştir. Diğer şekiller de daha hassas ve uzamsal olarak tutarlı sonuçlara işaret etmektedir.



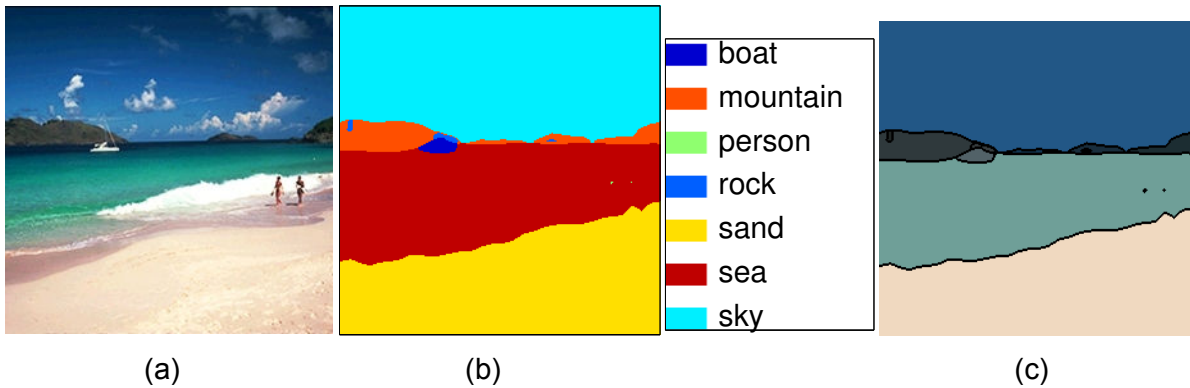
Şekil 4.2. Görsel karşılaştırma: (a) Orijinal görüntü; (b) Doğru etiketleme; (c) SuperParsing; (d) DC3.

## 4.2.1 Derin Ağlarda Bağlamsal Çıkarsama

Makine öğrenmede derin ağların geliştirilmesi ile birlikte, evrimsel sinir ağı (CNN) mimarisinin kullanılması görüntülerin piksel seviyesinde etiketlenerek yüksek doğruluk oranları elde edilmesini sağlamıştır (Shelhamer vd., 2016; Liu vd., 2016). Evrimsel ağlar çok sayıda birleştirme (pooling) ve üst-örnekleme (upsample) katmanı kullanarak uzamsal olarak yoğun fakat yumuşak sınıflandırma sunmaktadır.

Shelhamer vd. (2016)'daki tam evrimsel ağ mimarisi (Fully Convolutional Network (FCN)) farklı çözünürlükteki bağlamsal bilginin kaynaştırılması için kaba katman öznelikleri ince, alt-katman öznelikle birleştirmektedir. Ne var ki, ağ katmanları dikdörtgüsel evrim ve düzenli örnekleme ile üretildiği için, bu katmanlardaki öznelıklar sahnedeki nesnelere geometrisi ile uyumamaktadır.

Bu bölümde sunulan çalışmada, geliştirdiğimiz MRA modeli, FCN bölütleme sonuçlarının iyileştirilmesi için FCN ağının çıktı katmanında bir art-işlem adımı olarak kullanılmıştır. FCN mimarisi çıkış katmanında piksel seviyesinde sınıf skorları ve etiketler sağlamaktadır. Ters evrim katmanlarından dolayı bu skorlar komşu pikseller arasında yumuşak değişmektedir. FCN bölütleme sonucunu süperpikselle ayırmak için, FCN çıktısının bağlantılı bileşenlerini belirliyoruz. Bunun için aynı sınıf etiketine sahip komşu pikseller aynı bölüt/bileşene atanır (bkz Şekil 4.3). Daha sonra bölüt skoru piksellerinin skorlarının ortalamasına eşitlenir. Bu skorlar, SuperParsing'de olduğu gibi, çok-hipotezli MRA modelinde veri masraflarının tanımlanmasında kullanılır.



Şekil 4.3. FCN bölütleme: (a) Orijinal görüntü; (b) FCN etiketleri; (c) FCN süperpikselleri.

Bunun dışında, SuperParsing için, Bölüm 3.5'te anlatıldığı gibi, CNN özneliklerini geri erişim kümesi belirlemede global eşleştirme için ve yerel süperpiksel eşleştirme için kullanıyoruz. Bunlar sinir ağının en son sınıflandırma katmanından önceki katmandan çıkartılan öğrenilmiş özneliklerdir. VGG-F (Chatfield vd., 2014) ve AlexNet (Krizhevsky vd., 2012) ağlarından öğrenilmiş CNN öznelikleri kullanılmıştır. Bu derin ağlar ILSVRC

ImageNet veri setinde (Deng vd., 2009) eğitilmiştir. Her süperpiksel için CNN özniteliği, tüm görüntünün süperpiksel bölgesi hariç sifra atanması ve ağ çıktısının hesaplanması ile bulunur.

### Benzetimler

Benzetimlerde çok-hipotezli MRA yöntemi şu iki alternatif bölütleme sonucuna uygulanmıştır:

- $SP_1$ :  $(K, S, R) = (200, 100, 30)$ . SIFT öznitelikleri için KCB ya da LLC kodlama. Global ve süperpiksel eşleştirme için VGG-F/AlexNet öznitelikleri.
- $SP_2$ : FCN-8s (Shelhamer vd., 2016) derin mimarisi ile bölütleme sonucu.

Tablo 4.3'te SIFT Flow veri kümesi için DC1 ve DC2 modellerinin piksel-seviyesi ve ortalama sınıf-seviyesi doğruluk yüzdeleri sunulmuştur. Tabloda ayrıca karşılaştırma amacıyla, bölütleme-arası komşuluklar olmadan,  $SP_1$  ve  $SP_2$  sonuçlarına da yer verilmiştir (yani  $\lambda_n^m = 0 \quad \forall (n, m) \in IC$ ). Ayrıca SuperParsing, orijinal FCN-8s ve bazı diğer süperpiksel-tabanlı ve CNN-tabanlı ayırıştırma yöntemlerinin SIFT Flow veri kümesindeki performanslarına da yer verilmiştir.

Tablo 4.3. SIFT Flow veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Piksel Doğruluğu (%)	Sınıf-ortalaması (%)
DC1	86.8	50.8
DC2	<b>87.2</b>	50.4
$SP_1$	80.9	37.2
$SP_2$	86.0	51.6
SuperParsing	76.2	29.1
FCN-8s	85.9	53.9
Ates ve Sunetci (2017b)	80.6	31.8
George (2015)	81.7	50.1
Nguyen vd. (2016)	83.3	49.4
Liu vd. (2016)	86.8	52.0
Cheng vd. (2017)	86.4	49.4
Shuai vd. (2016a)	85.3	<b>55.7</b>

DC2 modeli önerilen çok-hipotezli MRA modelini kullanarak FCN-8s'in piksel doğruluğunu %1.3 artırmaktadır. DC1 doğruluğu DC2'ye göre %0.4 daha düşüktür. SIFT Flow veri seti için bu sonuç, bir kez daha, DC2'nin farklı veri masrafları ve bağlamsal kısıtları olan alternatifler arasından tercih yapmada daha esnek bir gösterim sunduğunu

kanıtlamaktadır. Dikkat edilirse, bölütleme-arası komşuluklar olmadan, MRA yumuşatması ile  $SP_2$  FCN-8s'e göre sadece %0.1'lik iyileşme sağlayabilmektedir. Dolayısıyla önerdiğimiz MRA modeli gerçekten de alternatif yöntemlerin bölütleme kararlarını başarıyla kaynaştırabilmektedir. Bölüm 4.2'deki sonuçlarla karşılaştırıldığında, hem CNN özniteliklerinin hem de FCN bölütlemenin kullanılıyor olması ayrıştırma doğruluğunda çok ciddi artış sağlamıştır.

Hem DC2 hem de DC1'in ortalama sınıf doğrulukları FCN-8s'ten düşüktür. Bunun bir nedeni MRA eniyilemenin etiketleri uzamsal olarak yumuşatması; diğer nedeni ise  $SP_1$ 'in ortalama sınıf doğruluğunun düşük olması ve dolayısıyla genel performansa katkı yapmamasıdır. Bölüm 4.2'de her iki hipotezin benzer performans gösterdiği durumda MRA modelinin ortalama sınıf doğruluklarını da iyileştirdiğini görmüştük. Dolayısıyla eğer  $SP_1$ 'e daha başarılı bir süperpiksel-tabanlı ayrıştırma algoritması atanırsa (ör., George (2015)), FCN-8s'in sınıf ortalamasının da iyileşmesini bekliyoruz. Aslında, Tablo 4.3'ten görüldüğü üzere, düşük başarılı bir  $SP_1$  alternatifi ile dahi piksel doğruluğunun en son süperpiksel-tabanlı ve CNN-tabanlı bölütleme yöntemlerinin üstünde çıkması MRA modelinin potansiyelini ortaya koymaktadır.

Tablo 4.4. LabelMe veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

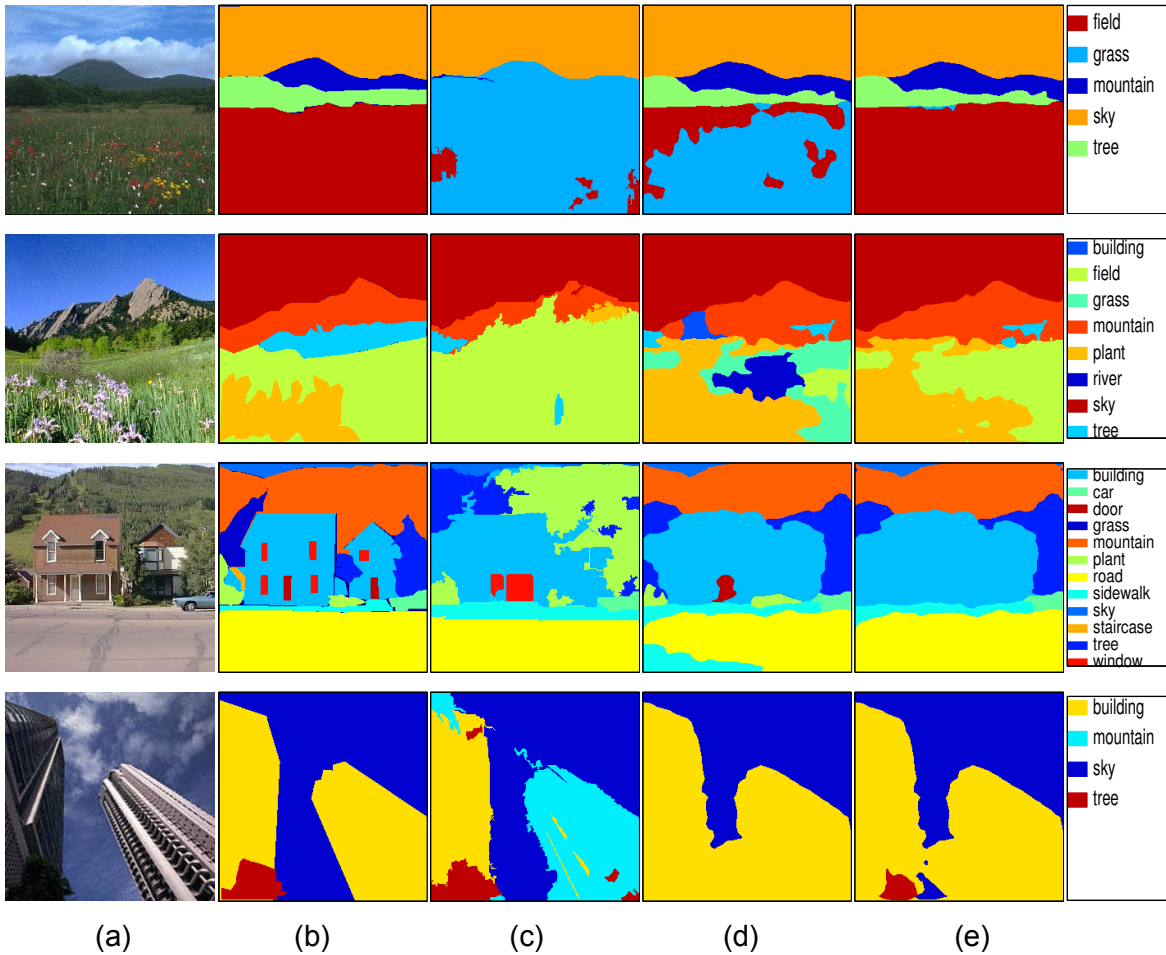
Yöntem	Piksel Doğruluğu	Sınıf-ortalaması
DC1	<b>90.4</b>	75.5
DC2	89.9	<b>76.1</b>
$SP_1$	86.7	67.3
$SP_2$	89.0	72.5
FCN-8s	86.9	70.0
Ates ve Sunetci (2017b)	83.8	55.6
Myeong ve Lee (2013)	81.8	54.4
Nguyen vd. (2015)	82.7	55.1
Nguyen vd. (2016)	85.5	63.2

Tablo 4.4'te LabelMe veri kümesi için test edilen yöntemlerin piksel-seviyesi ve ortalama sınıf-seviyesi doğruluk yüzdeleri sunulmuştur. Literatürden karşılaştırma için sunulan sonuçlar Nguyen vd. (2015, 2016), Myeong ve Lee (2013), ve Bölüm 4.2'deki sonucumuzdur. Bu veri kümesinde, önerdiğimiz yöntem varolan en iyi sonucu piksel doğruluğunda %4.9, sınıf doğruluğunda ise %12.9 geçmiştir. Ayrıca FCN-8s sonucunu da piksel ve sınıf doğruluğunda, sırasıyla %3.5 ve %6.1 iyileştiriyoruz.

LabelMe veri kümesi için DC1 sonucu DC2'ye göre piksel doğruluğunda %0.5 daha iyidir, fakat DC2 daha yüksek ortalama sınıf doğruluğu sunmaktadır. DC1 sonucu  $SP_2$ 'nin

piksel ve sınıf doğruluğunu, sırasıyla, %1.4 ve %3.0 iyileştirmiştir. Bu sonuç bir kez daha bölütleme-arası komşulukların alternatif yöntemlerdeki tümler bilgiyi birleştirmedeki önemini ortaya koymaktadır. SIFT Flow'dakinin aksine, algoritmamız bu veri kümesinde ortalama sınıf doğruluklarında da artış sağlamıştır. Dolayısıyla,  $SP_1$  ve  $SP_2$ 'nin benzer performanslara sahip oldukları durumda, çok-hipotezli MRA yaklaşımı test edilen yöntemlerin sadece piksel doğruluğunu değil ortalama sınıf doğruluklarını da artırmaktadır.

Şekil 4.4'te SIFT Flow'dan bazı test görüntüleri için  $SP_1$ ,  $SP_2$  ve DC2 ayrıştırma sonuçları görsel olarak karşılaştırılmıştır. DC2 etiketleri genel olarak  $SP_1$  ve  $SP_2$ 'ye göre daha tutarlı ve doğrudur. En üst şekilde her iki yöntemin *çimen* olarak etiketlediği bölge *yeşil alan* olarak doğru tanımlanmıştır. Diğer üç şekilde ve genelde tüm SIFT Flow test kümesinde DC2 sonucu en az  $SP_1$  ve  $SP_2$  sonuçlarının iyisi kadar veya daha iyidir. Başka deyişle önerilen yaklaşım, bölütleme-içi ve arası komşuluklardaki karmaşık bağlamsal kısıtları kullanarak, iki hipotezden daha olası etiketi doğru şekilde seçmeyi başarmaktadır.



Şekil 4.4. Görsel karşılaştırma: (a) Orijinal görüntü; (b) Doğru etiketleme; (c)  $SP_1$ ; (d)  $SP_2$ ; (e) DC2.

## 5. UZAKTAN ALGILAMA İÇİN GÖRÜNTÜ AYRIŞTIRMA

Projenin son döneminde, geliştirilen yöntemler uzaktan algılama uygulamaları için test edilmiştir. Uydu ya da uçaktan çekilen görüntülerin bölütlenmesi ve anlambilimsel sınıflara ayrılması, arazi örtüsü ya da arazi kullanımı sınıflandırma, farklı nesnelerin (bina, araba, yol, vs.) tespiti ve etiketlenmesi gibi çeşitli problemler kapsamında ele alınabilir. Biz bu projede yerleşim bölgelerinin yüksek çözünürlüklü çoklu-spektral optik görüntüleri ve yeryüzü şekillerinin yükseklik bilgisini kullanarak nesne etiketleme problemini inceledik.

Süperpiksellerin uzaktan algılama problemlerinde kullanılmasında çok sayıda çalışmaya rastlanılmaktadır. Arazi örtüsü sınıflandırma (Kluckner vd., 2010) ve değişiklik sezimi problemlerinde (Wu vd., 2012; Gadhiraaju vd., 2014) süperpiksel kullanan çalışmalar bulunmaktadır. Fakat bu çalışmalar genelde standart süperpiksel bölütleme ve güdümlü/güdümsüz sınıflandırma yöntemlerinin ilgili uzaktan algılama problemine uyarlanmasının ötesine geçmemektedir. Stefanski vd. (2013) arazi örtüsü etiketleme için süperpiksel bölütleme parametrelerinin çapraz-sağlama ile eniyilenmesini önermişlerdir. Fakat bu makalede tek bir süperpiksel yöntemi üzerinde durulmaktadır ve spektral değerler dışında herhangi bir betimleyici denenmemiştir.

Süperpiksel-tabanlı bölütleme ve ayırıştırma için önerdiğimiz benzer bir yaklaşım, ISPRS (International Society for Photogrammetry and Remote Sensing) tarafında yürütülen 2-Boyutlu Anlambilimsel Etiketleme Yarışması (ISPRS, 2017) için denektaşı (benchmark) yöntem olarak önerilen Gerke'nin çalışmasıdır (Gerke, 2014). Bu çalışmada süperpikseller için farklı özniteliklerle birlikte düzgelenmiş yükseklik verisi ve düzgelenmiş bitki örtüsü indisi de kullanılarak Adaboost-tabanlı sınıflandırıcı ile eğitim yapılmaktadır. Aynı yarışma için sunulan diğer yöntemlerde standart özniteliklerin yanında, CNN özniteliklerinin de kullanıldığı ve derin CNN modellerinin eğitildiği gözlenmektedir ((Audebert vd., 2017), (Marmanis vd., 2016). Eğitimde yükseklik verisinin ve bitki örtüsü indisinin de spektral bantlarla birlikte kullanılmasının sınıflandırma doğruluğunu artırdığı gösterilmiştir.

Bu bölümde projede geliştirdiğimiz yöntemlerin ISPRS'in 2-Boyutlu Anlambilimsel Etiketleme Yarışması için sunduğu eğitim veri kümesine uygulanması sonucu elde edilen sınıflandırma sonuçları tartışılmaktadır. Bölüm 5.1'de veri kümesi teknik özellikleri açıklandıktan sonra Bölüm 5.2'de kullanılan yaklaşım, farklı öznitelikler ve benzetim sonuçları karşılaştırmalı olarak sunulmaktadır.

## 5.1 ISPRS 2-Boyutlu Anlambilimsel Etiketleme Yarışması

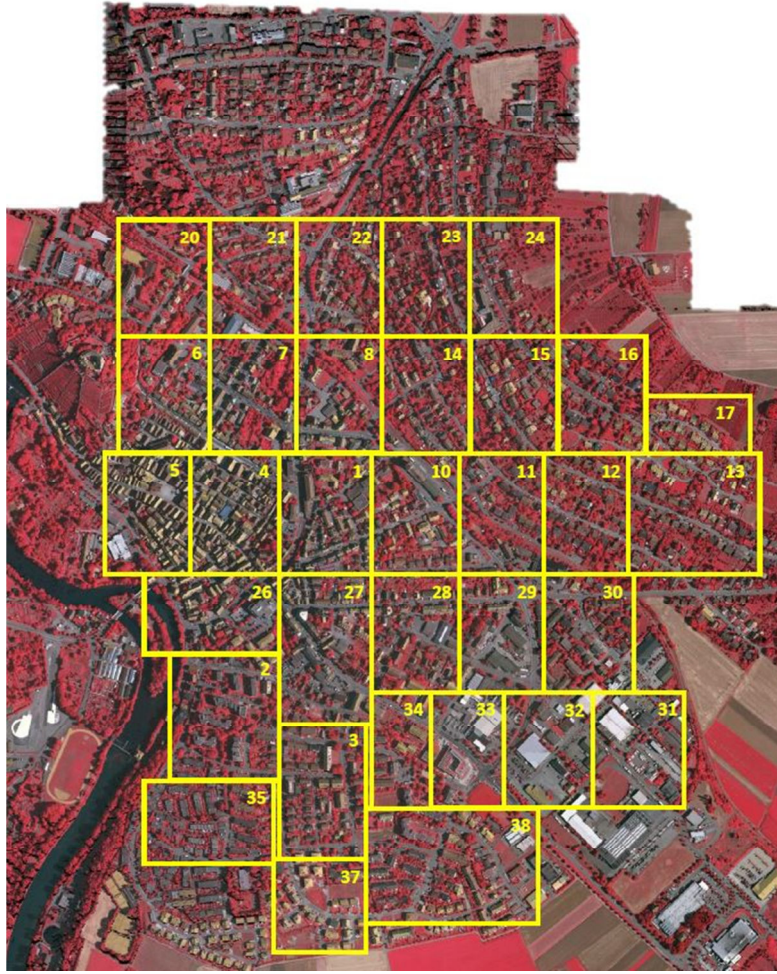
Uzaktan algılama çözümlerini zorlaştıran en önemli sorun, görüntü içeriklerinin kamera özellikleri (çözünürlük, çalıştığı spektral bantlar, vs.) ve mevsimsel ve atmosferik koşullara bağlı olarak büyük farklılıklar göstermesidir. Bu nedenle uzaktan algılama veri kümeleri, belli bir kamerayla belli bir zamanda çekilen ve geniş bir alanı kapsayan görüntülerin eğitim ve test veri kümesi olarak farklı kısımlara ayrılması yoluyla oluşturulmaktadır.

ISPRS 2-Boyutlu Anlambilimsel Etiketleme Yarışması, eğitim/sağlama ve test amaçlı olarak iki farklı veri kümesini araştırmacıların kullanımına sunmaktadır. Vaihingen/Almanya ve Potsdam/Almanya şehirlerinin havadan çok yüksek çözünürlüklü gerçek orto fotoğraflarından oluşan bu veri kümelerinde ayrıca yükseklik bilgisi için dijital yüzey modelleri (Digital Surface Model (DSM)) de yer almaktadır. Biz proje kapsamında Vaihingen veri kümesini kullandık (Bkz. Şekil 5.1).

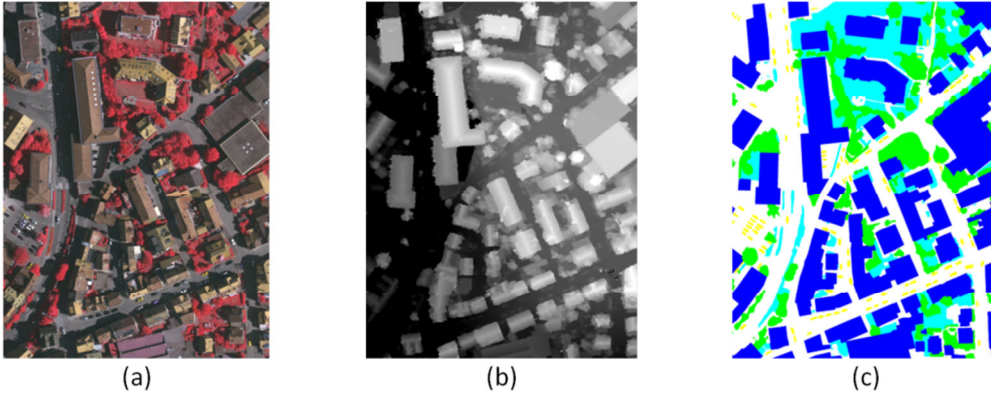
Vaihingen veri kümesinde büyük bir mozaik görüntüden kesilmiş farklı boyutlarda 33 ortofoto yer almaktadır. Ayrıca fotoğraflarla hizalanmış DSM verisi gri-seviyeli görüntüler şeklinde sunulmuştur (bkz. Şekil 5.2(b)). Ortofoto ve DSM görüntülerinin yer örnekleme uzaklığı 9 cm'dir. Ortofotolar 8-bitlik 3-bant multi-spektral görüntülerdir: kullanılan spektral bantlar, yakın infrared (IR), kırmızı (R) ve yeşil (G) bantlarıdır (bkz. Şekil 5.2(a)). DSM yükseklikleri 32-bit kayan noktalı değerlerdir. Vaihingen'deki 33 görüntüden 16'sı eğitim/sağlama için, geri kalan 17 tanesi test görüntüsü olarak ayrılmıştır. Görüntülerin 6 sınıfa bölütlenmesi amaçlanmaktadır:

1. Geçirgen olmayan yüzey (yol, kaldırım, vs.)
2. Bina
3. Alçak bitki örtüsü
4. Ağaç
5. Araba
6. Gürültü/arka-plan

Gürültü/arka-plan sınıfında diğer sınıflara girmeyen diğer nesnelere (ör. havuz, nehir, tenis sahası, vs.) yer almaktadır. Vaihingen eğitim kümesi 16 görüntünün referans doğru etiketleme bilgisini içermekle birlikte test kümesindeki görüntüler için doğru etiketleme paylaşılmamıştır. Araştırma ekipleri eğitim kümesi üzerinde geliştirdikleri algoritmalarını yarışmaya göndererek test verisinde değerlendirilmesini ve etiketleme doğruluğunun tespitini talep edebilmektedirler. Algoritma performansları web sayfası üzerinden de paylaşılmaktadır (ISPRS, 2017).



Şekil 5.1. Vaihingen ortofoto mozaïği.



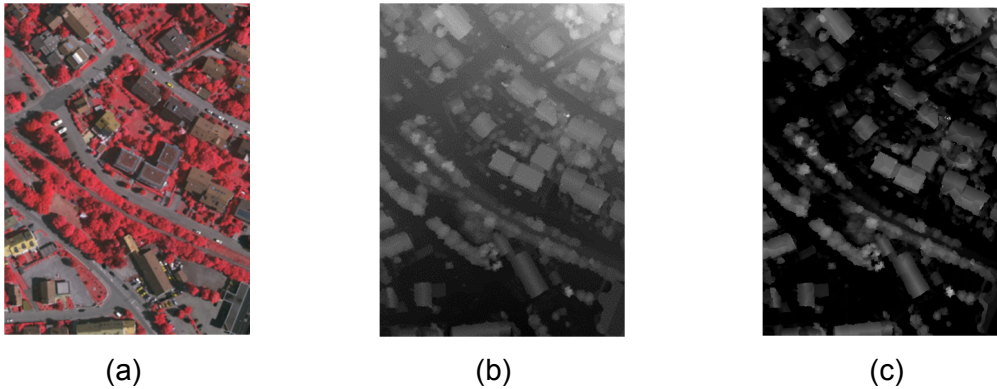
Şekil 5.2. Vaihingen eğitim verisi: (a) 3-bant multispektral görüntü; (b) DSM görüntüsü; (c) Doğru etiketleme.



## 5.2 Benzetimler

Geliştirdiğimiz süperpiksel-tabanlı etiketleme yöntemleri ve çok-hipotezli MRA modelinin Vaihingen veri kümesinde test edilmesi için, doğru etiket bilgisi varolan 16 görüntüyü kullandık. Diğer algoritmalarla karşılaştırabilmek için geliştirilen algoritmaların yarışmaya gönderilmesi proje sonrasında yapılması düşünülen çalışmalardan birisidir. 16 görüntüyü, 10 eğitim ve 6 sağlama görüntüsü olarak ayırdık. Bu bölümde sunulan sonuçlar bu 6 sağlama görüntüsünden elde edilen etiketleme doğruluklarına karşılık gelmektedir. Görüntüler Vaihingen şehrinin farklı bölgelerine denk geldiği ve benzer içerikler barındırdığı için algoritmalarımızın 17 test görüntüsü üzerinde de benzer sonuçlar üreteceğini beklemekteyiz.

SuperParsing algoritmasında önceki bölümlerde bahsedilen öznelikler yanında DSM yükseklik verisi de bir öznelik olarak kullanılmıştır. Fakat yeryüzü eğimi dikkate alındığında DSM verisinin ayırt edici olabilmesi için nesnelerin yerden yükseliğine ihtiyaç vardır. Bunun için Gerke (2014) tarafından üretilen düzgelenmiş DSM (NDSM) verisi kullanılmıştır (bkz Şekil 5.3). NDSM’de yükseklik  $[0,255]$  aralığında düzgelenmiş ve 8-bit gri-seviye görüntü olarak kaydedilmiştir. Bu veride yolların yüksekliği sıfırdır. Her süperpiksele ait piksellerin NDSM yüksekliklerinin ortalaması, süperpiksel için yeni bir öznelik betimleyici olarak kullanılmıştır.



Şekil 5.3. Vaihingen yükseklik verisi: (a) 3-bant multispektral görüntü; (b) DSM görüntüsü; (c) NDSM (Düzgelenmiş yükseklik) görüntüsü.

FCN-8s mimarisi Vaihingen veri kümesine uyarlanmış ve 10 görüntü ile eğitilmiştir. Fakat Vaihingen görüntülerinin boyutları çok büyük olduğu için bu derin mimariye uygun değildir. Bu nedenle görüntüler 256x256’lık alt-görüntülere bölünmüş, eğitim ve test süreçleri bu alt-görüntülerle yürütülmüştür. Alt-görüntüler sınırlarda 64-piksel örtüşecek şekilde tanımlanmış; örtüşen pikseller için sınıf skorları farklı alt-görüntülerden gelen skorların ortalaması olarak atanmıştır.

Benzetimler için iki farklı FCN-8s mimarisi eğitilmiştir. Birincisinde (FCN1) 3-bantlı spektral (IR,R,G) görüntüler kullanılmıştır. İkinci mimari için ise NDSM verisi ve düzgelenmiş bitki örtüsü indisi (Normalized Digital Vegetation Index (NDVI)) kullanılmıştır ( $p$  pikseli için):

$$DVI(p) = \frac{IR(p) - R(p)}{IR(p) + R(p)} \quad (5.1)$$

DVI değeri [0 255] aralığına düzgelendikten sonra, FCN2 mimarisi (NDSM,NDVI,G) bantları kullanılarak elde edilen 3-bantlı görüntüler ile eğitilmiştir.

Benzetimlerde çok-hipotezli MRA yöntemi şu alternatif bölütleme sonuçlarına uygulanmıştır (kısaltmalar için bkz. Bölüm 3.5):

- $SP_1$ : SuperParsing ayrıştırma,  $(K, S, R) = (5, 200, 15)$ . EK: 5. SB: Kısıtlı, KCB, NDSM( $w_k = 3$ ).
- $SP_2$ : Süperpikseller için BDT sınıflandırıcı,  $(K, S) = (5, 200)$ . SB: Kısıtlı, KCB, NDSM( $w_k = 3$ ).
- $SP_3$ : FCN1 skorları ile bölütleme sonucu.
- $SP_4$ : FCN1 + FCN2 skorları ile bölütleme sonucu.

Tablo 5.1'de Vaihingen'den seçilen 6 doğrulama görüntüsü için,  $SP_1$ ,  $SP_2$ ,  $SP_3$ ,  $SP_4$  ve bu alternatiflerden bazıları için DC1, DC2 modellerinin piksel-seviyesi ve ortalama sınıf-seviyesi doğruluk yüzdeleri sunulmuştur.  $SP_2$  sonucu,  $SP_1$  ile aynı süperpiksel ve öznitelikler için SuperParsing algoritması yerine BDT sınıflandırıcı kullanılması yoluyla elde edilmiştir. Vaihingen eğitim kümesinde sadece 10 görüntü bulunması nedeniyle SuperParsing'in başarımı düşük çıkmaktadır. BDT sınıflandırıcı kullanılması sayesinde piksel doğruluğu %2.4, sınıf ortalaması %5.8 artmıştır. FCN1 bölütleme sonucu, sınıf ortalaması yüksek olmasına karşın, piksel doğruluğunda düşük kalmıştır. Bunun temel nedeni de NDSM verisinin kullanılmamış olmasıdır.  $SP_4$  için FCN1 ve FCN2 skorlarının toplamı kullanılmıştır. Böylelikle NDSM ve NDVI bilgisinin kaynaştırılması sonucu piksel doğruluğu %4.6, sınıf ortalaması %1.5 artmıştır.

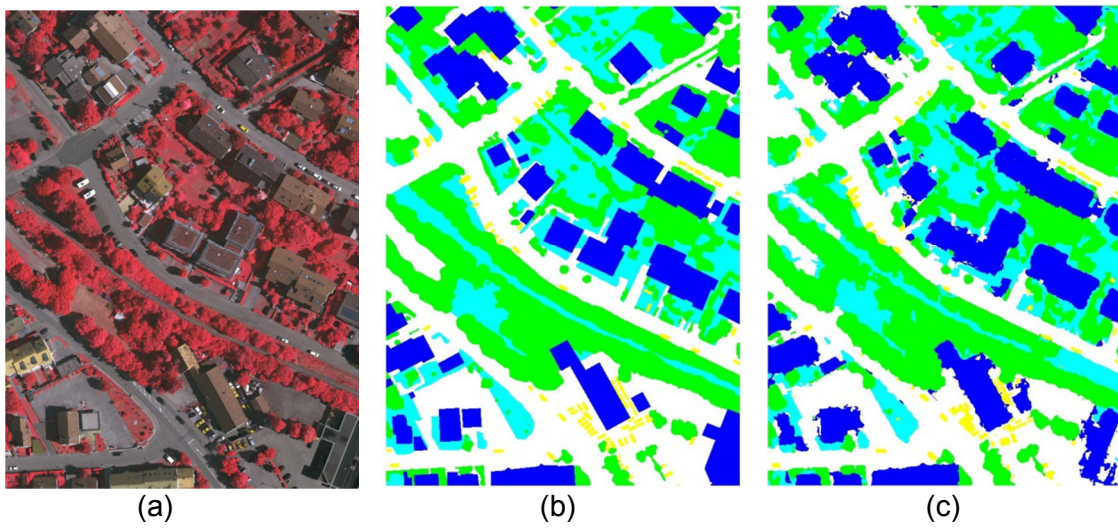
$SP_2$  ve  $SP_3$  için çok-hipotezli MRA'da DC1 modelinin kullanılması ile piksel doğruluğu %83.5 olmuştur.  $SP_2$  ve  $SP_4$  için ise DC1 modeli %84.7 doğruluk vermiştir. Dolayısıyla çok-hipotezli MRA'da  $SP_3$  yerine  $SP_4$  kullanılmasının sağladığı doğruluk artışı %1.2'de kalmıştır. Bunun temel nedeni  $SP_2$ 'de NDSM bilgisinin zaten kullanılıyor olması ve dolayısıyla  $SP_4$  ile kaynaştırılmasının sağladığı tümler bilginin daha az etkili olmasıdır.  $SP_2$  ve  $SP_4$  için DC2 modeli piksel doğruluğunda DC1'den çok az düşük çıkmakla birlikte sınıf ortalamasında %80.2 doğruluk ile en iyi sonucu vermiştir.

Tabloda ayrıca ISPRS (2017)'de yarışan bazı algoritmaların 17 görüntülü test verisi üzerinde sağladıkları sonuçlar da karşılaştırma amacıyla listelenmiştir. Farklı görüntülerle elde edilen sonuçlar olduğu için tam sağlıklı bir karşılaştırma yapmak mümkün olmamakla birlikte çok-hipotezli MRA algoritmasının diğer yöntemler arasında ortalamanın üstünde iyi sonuçlar verdiği gözlenmektedir. Proje sonrası devam edecek geliştirme çalışmalarıyla en iyi etiketleme doğruluklarının da üstüne çıkılması hedeflenmektedir.

Tablo 5.1. Vaihingen veri kümesinde etiketleme doğruluk (piksel/sınıf ort.) karşılaştırması

Yöntem	Piksel Doğruluğu (%)	Sınıf-ortalaması (%)
$SP_1$	78.9	65.4
$SP_2$	81.3	71.2
$SP_3$	78.0	76.7
$SP_4$	82.6	78.2
DC1 ( $SP_2, SP_3$ )	83.5	78.9
DC2 ( $SP_2, SP_3$ )	83.4	78.6
DC1 ( $SP_2, SP_4$ )	84.7	79.5
DC2 ( $SP_2, SP_4$ )	84.5	80.2
Gerke (2014)	81.6	74.7
Marmanis vd. (2016)	85.6	80.6
Audebert vd. (2017)	<b>86.9</b>	<b>81.3</b>

Şekil 5.4'te bir doğrulama görüntüsü için DC2 ( $SP_2, SP_4$ )'nin etiketleme sonucu sunulmuştur. Ayrıştırma başarımı genel olarak iyi olmakla birlikte, özellikle araba sınıfı için sonuçların iyileştirilmesinin mümkün olduğu gözükmektedir.



Şekil 5.4. Görsel sonuç (beyaz: yüzey, mavi:bitki örtüsü, lacivert: bina, yeşil: ağaç, sarı: araba): (a) 3-bant multispektral görüntü; (b) Doğru etiketleme; (c) DC2 ( $SP_2, SP_4$ ) sonucu.

## 6. SONUÇ VE ÖNERİLER

Bu projede açık hava sahneleri ve uzaktan algılama görüntülerinde anlambilimsel bölütleme için özgün süperpiksel-tabanlı yöntemler geliştirilmiştir. Özellikle sınıflandırıcı eğitimi gerektirmeyen veri güdümlü SuperParsing yaklaşımı üzerinde durulmuş ve bu algoritmanın farklı bileşenleri iyileştirilerek etiketleme başarımı artırılmıştır. Global ve süperpiksel betimleyici seçimi, kodlanması, olabilirlik oran modellenmesi ve kestirimi, çok-hipotezli bölütleme ve bağlamsal modelleme konularında yapılan çalışmalar ile ayrıştırma başarımında kayda değer iyileşme sağlanmıştır.

Son birkaç yıl içerisinde derin evrimsel ağlarla (CNN) sınıflandırma konusunda yaşanan gelişmeler paralelinde anlambilimsel bölütleme için derin ağların kullanılması popüler bir araştırma alanı olmuş ve çok ciddi performans artışları sağlanmıştır. Projede yaptığımız çalışmaların güncel kalabilmesi adına, proje iş planında olmadığı halde, geliştirdiğimiz yöntemlerde derin ağların kullanılması için çeşitli çözümler geliştirilmiştir. Global ve süperpiksel eşleştirmede “el yapımı” öznitelikler yanında derin sinir ağlarında öğrenilmiş özniteliklerin de kullanılması bu çözümlerden birisidir. Diğer başarılı yaklaşım ise çok-hipotezli bağlamsal modelin süperpiksel-tabanlı ayrıştırma sonuçları ile derin ağlarla elde edilmiş bölütleme sonuçlarının kaynaştırılması için kullanılması olmuştur. Farklı yöntemlerle elde edilen bölütleme sonuçlarının kaynaştırılması için uyarlanabilir, esnek ve etkili olduğu ortaya konulan çok-hipotezli MRA modeli ile literatürde en son elde edilen ayrıştırma doğruluklarının üstüne çıkılmıştır.

Proje çıktıları iki yurtdışı, bir yurtiçi konferansta bildiri olarak sunulmuştur. Çok-hipotezli MRA modeli ile ilgili bir dergi makalesi hazırlanmış ve Elsevier Pattern Recognition Letters dergisine gönderilmiştir. Proje lisans bursiyeri Elçin Değirmenci bitirme tezinde bağlamsal modelleme konusunda yaptığı benzetimlere yer vermiştir. Proje yüksek lisans bursiyeri Sercan Süneci yüksek lisans tezini öznitelik kodlama ve CNN özniteliklerinin kullanımı ile ilgili yaptığı çalışmalardan hazırlamıştır ve proje sonrası savunmasını yapacaktır.

Proje tamamlandıktan sonraki süreçte derin ağlarla ilgili başlatılan çalışmaların geliştirilmesi düşünülmektedir. Proje süresince ortaya çıkan, ileriye dönük yapılması düşünülen çalışmalar şöyle özetlenebilir:

- Şu an için çok-hipotezli MRA modeli, eğitilmiş evrimsel ağın çıktısına bir art-ışlem olarak uygulanmaktadır. İleride bağlamsal modellemenin derin ağ mimarisine entegre edilmesi için çalışılacaktır. Böylelikle tüm algoritmanın baştan

sona eđitilmesi m¼mk¼n olacaktır. Çok-hipotezli MRA modeli parametrelerinin bu eđitim sırasında ¼đrenilmesinin ayrıştırma dođruluđu ¼zerindeki ¼nemli bir etkisi olacađı tahmin edilmektedir.

- Çok-hipotezli yaklařımda farklı s¼perpiksel-tabanlı ve CNN-tabanlı y¼ntemlerin kaynařtırılması denenecektir. İki den fazla alternatif kullanılmasının sađlayacađı kazanç analiz edilecektir. S¼perpiksel-tabanlı ve CNN-tabanlı b¼l¼tleme y¼ntemlerinin ortak bir mimaride nasıl birlikte eđitilip eniyilenebilecekleri arařtırılacaktır.
- Uzaktan algılama i¼in yapılan ¼alıřmalar s¼rd¼r¼lecektir. Burada kullanılan veri k¼mesine uygun CNN mimarilerinin geliřtirilmesine ihtiya¼ vardır. Ayrıca farklı betimleyicilerle eđitilmiř birden fazla mimarinin nasıl kaynařtırılabileceđi de ilgin¼ bir arařtırma problemidir.
- Geliřtirilen ¼ok-hipotezli bađlamsal model bilgisayarla g¼r¼ ve uzaktan algılama alanlarında bařka problemlere de uygulanacaktır. Nesne tanıma ve sezimi, deđiřiklik sezimi, afet hasar tespiti gibi problemler i¼in bu projede geliřtirilen y¼ntemlerin uyarlanması ¼zerinde durulacaktır.

G¼n¼m¼zde yapay zeka alanında yapılan arařtırma ve geliřtirme ¼alıřmaları b¼y¼k hız kazanmıř durumdadır. Bilgisayarla g¼r¼n¼n pek ¼ok alanındaki problemler i¼in her g¼n yeni ve daha iyi sonu¼lar veren ¼z¼mler ¼retilmektedir. Bu projede sahne i¼eriđi anlamlandırmada gerekli olan b¼l¼tleme ve etiketleme i¼in farklı yaklařımların eniyi kaynařtırılması ile daha iyi sonu¼lar elde edilebileceđi ortaya konmuřtur. ¼n¼m¼zdeki d¼nemde b¼l¼tleme, etiketleme, bađlamsal modelleme problemlerini ortak bir bakıř a¼ısı altında ele alan ¼zg¼n derin sinir ađı mimarileri geliřtirilmesi ve eđitilmesi ile bug¼nk¼ sonu¼ların ¼ok daha ¼st¼nde performanslar elde edilmesi m¼mk¼n g¼z¼kmektedir. Bu ve benzeri ¼alıřmalara proje sonrasında devam edilmesi ve proje boyunca elde edilen bilgi birikiminin yeni arařtırma projelerinde deđerlendirilmesi hedeflenmektedir.

## REFERANSLAR

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. 2012. "SLIC superpixels compared to state-of-the-art superpixel methods", IEEE Trans. Pattern Anal. Mach. Intell., 34 (11), 2274-2282.
- Ak, K.E., Ateş, H.F. 2015. "Çok-hipotezli süperpikseller ile sahne bölütleme ve etiketleme", IEEE Sinyal İşleme İletişim Uyg. Kurultayı (SİU 2015).
- Aksoy, S. 2006. "Spatial techniques for image classification", 491-513. Signal and Image Processing for Remote Sensing. Editör: Chen, C.H. Boca Raton: CRC Press.
- Ates, H.F.**, Sunetci, S., Ak, K.E. 2016. "Kernel likelihood estimation for superpixel image parsing", Proc. Int. Conf. Image Analysis Recog. (ICIAR 2016), 234-242.
- Ateş, H.F.**, Süneci, S. 2017(a). "Kernel kod-tablosu kodlaması ile sahne etiketleme", IEEE Sinyal İşleme İletişim Uyg. Kurultayı (SİU 2017).
- Ates, H.F.**, Sunetci, S. 2017(b). "Improving semantic segmentation with generalized models of local context", Proc. Int. Conf. Computer Analysis Images Patterns (CAIP 2017), 320–330.
- Ates, H.F.**, Sunetci, S. 2017(c). "Multi-hypothesis contextual modeling for semantic segmentation", Pattern Recog. Letters, *gönderildi*.
- Audebert, N., Le Saux, B., Lefevre, S. 2016. "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks", Asian Conf. Computer Vision (ACCV 2016), 180-196.
- Boykov, Y., Kolmogorov, V. 2004. "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision", IEEE Trans. Pattern Anal. Mach. Intell, 26 (9), 1124-1137.
- Budnik, M, Gutierrez-Gomez, E.L., Safadi, B., Quénot, G. 2015. "Learned features versus engineered features for semantic video indexing", Proc. 13th Int. Workshop Content-Based Multimedia Indexing (CBMI), 1-6.
- Chatfield, K., Lempitsky, V., Vedaldi, A., Zisserman, A. 2012. "The devil is in the details: an evaluation of recent feature encoding methods", Proc. British Machine Vision Conf., 76.1-76.12.
- Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A. 2014. "Return of the devil in the details: delving deep into convolutional nets", Proc. British Machine Vision Conf.
- Cheng, F., He, X., Zhang, H. 2017. "Stacked learning to search for scene labeling", IEEE Trans. Image Process. 26 (4), 1887–1898.
- Deng, J., Dong, W., vd. 2009. "Imagenet: a large-scale hierarchical image database", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 248-255.
- Duin, R.P.W., 2002. "The combining classifier: to train or not to train?", Proc. 16th Int. Conf. Pattern Recog. (ICPR), 2, 765-770.
- Eigen, D., Fergus, R. 2012. "Nonparametric image parsing using adaptive neighbor sets", Proc. IEEE Conf. Computer Vision Pattern Recog. (CVPR), 2799-2806.

Farabet, C., Couprie, C., Najman, L., LeCun, Y. 2013. "Learning hierarchical features for scene labeling", IEEE Trans. Pattern Anal. Mach. Intell. 35(8), 1915–1929.

Felzenszwalb, P.F., Huttenlocher, D.P. 2004. "Efficient graph-based image segmentation", Int. Journal Computer Vision, 59 (2), 167-181.

Fulkerson, B., Vedaldi, A., Soatto, S. 2009. "Class segmentation and object localization with superpixel neighborhoods", Proc. IEEE 12th Int. Conf. Computer Vision (ICCV), 670-677.

Gadhiraju, S.V., Sahbi, H., Banerjee, B., Buddhiraju, K.M. 2014. "Supervised change detection in satellite imagery using super pixels and relevance feedback", GEOMATICA, 68 (1), 5-14.

van Gemert, J.C., Geusebroek, J.M., Veenman, C.J., Smeulders, A.W.M. 2008. "Kernel codebooks for scene categorization", Proc. European Conf. Computer Vision, 3, 696-709.

George, M. 2015. "Image parsing with a wide range of classes and scene-level context", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3622-3630.

Gerke, M. 2014. "Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (Vaihingen)", Teknik Rapor.

Gould, S., Fulton, R., Koller, D. 2009. "Decomposing a scene into geometric and semantically consistent regions", Proc. IEEE 12th Int. Conf. Computer Vision (ICCV), 1-8.

Hoiem, D., Efros, A., Hebert, M. 2007. "Recovering surface layout from an image", Int. Journal Computer Vision, 75, 151-172.

Huang, Y., Wu, Z., Wang, L., Tan, T. 2014. "Feature coding in image classification: a comprehensive study", IEEE Trans. Pattern Anal. Mach. Intell., 36 (3), 493-506.

ISPRS (International Society for Photogrammetry and Remote Sensing), "2D Semantic Labeling Contest", <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html> Son erişim tarihi: 25 Ekim 2017.

Jain, A., Gupta, A., Davis, L. 2010. "Learning what and how of contextual models for scene labeling", Proc. European Conf. Computer Vision, 199–212.

Kluckner, S., Donoser, M., Bischof, H. 2010. "Super-pixel class segmentation in large-scale aerial imagery", Proc. Annual Workshop Austrian Assoc. Pattern Recog., 131-138.

Krizhevsky, A., Sutskever, I., Hinton, G.E. 2012. "Imagenet classification with deep convolutional neural networks", Advances in Neural Info. Process. Systems, 25, 1097–1105.

Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K. 2009. "TurboPixels: fast superpixels using geometric flows", IEEE Trans. Pattern Anal. Mach. Intell., 31 (12), 2290-2297.

Liang, M., Hu, X., Zhang, B. 2015. "Convolutional neural networks with intralayer recurrent connections for scene labeling", Advances in Neural Info. Process. Systems, 937–945.

Liang-Chieh, C., Papandreou, G., *vd.* 2017. "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs", IEEE Trans. Pattern Anal. Mach. Intell.

- Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T. 2008. "SIFT Flow: dense correspondence across difference scenes", Proc. European Conf. Computer Vision (ECCV'08), 5304, 28-42.
- Liu, W., Rabinovich, A., Berg., A. 2016. "ParseNet: looking wider to see better", Proc. Int. Conf. Learning Represent. (ICLR) Workshop.
- Lowe, D.G. 1999. "Object recognition from local scale-invariant features", Proc. IEEE 7th Int. Conf. Computer Vision, 2, 1150-1157.
- Marmanis, D., Wegner, J.D., *vd.* 2016. "Semantic segmentation of aerial images with an ensemble of CNNs", ISPRS Annals Photogrammetry, Remote Sensing Spatial Info. Sciences, 473-480.
- Ming-Yu, L., Tuzel, O., Ramalingam, S., Chellappa, R. 2011. "Entropy rate superpixel segmentation", Proc. IEEE Conf. Computer Vision Pattern Recog. (CVPR), 2097-2104.
- Mostajabi, M., Yadollahpour, P., Shakhnarovich, G. 2015. "Feed-forward semantic segmentation with zoom-out features", Proc. *IEEE Conf. Comp. Vision Pattern Recog. (CVPR)*, 3376-3385.
- Myeong, H., Lee, K., 2013. "Tensor-based high-order semantic relation transfer for semantic scene segmentation", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3073-3080.
- Nguyen, T., Lu, C., Sepulveda, J., Yan, S. 2015. "Adaptive nonparametric image parsing", IEEE Trans. Circuits Syst. Video Tech., 25(10), 1565-1575.
- Nguyen, T.V., Liu, L., Nguyen, K. 2016. "Exploiting generic multi-level convolutional neural networks for scene understanding", Proc. 14th Int. Conf. Control, Automation, Robotics and Vision (ICARCV), 1-6.
- Oliva, A., Torralba, A. 2006. "Building the gist of a scene: The role of global image features in recognition", Progress in Brain Research, 155, 23-26.
- Park, E., Han, E., Berg, T.L., Berg, A.C. 2016. "Combining multiple sources of knowledge in deep CNNs for action recognition", Proc. IEEE Winter Conf. Appl. Computer Vision (WACV), 1-8.
- Perronnin, F., Sánchez, J., Mensink, T. 2010. "Improving the fisher kernel for large-scale image classification", Proc. European Conf. Computer Vision (ECCV), 6314, 143-156.
- Razzaghi, P., Samavi, S., 2014. "A new fast approach to nonparametric scene parsing", Pattern Recognition Letters, 42, 56-64.
- van de Sande, K.E.A., Gevers, T., Snoek, C.G.M. 2010. "Evaluating color descriptors for object and scene recognition", IEEE Trans. Pattern Anal. Mach. Intell., 32 (9), 1582-1596.
- Sharma, A., Tuzel, O., Liu, M. 2014. "Recursive context propagation network for semantic scene labeling", Advances in Neural Info. Process. Systems, 27, 2447-2455.
- Shelhamer, E., Long, J., Darrell, T. 2016. "Fully convolutional networks for semantic segmentation", IEEE Trans. Pattern Anal. Mach. Intell., 39 (4), 640-651.
- Shimazaki, H., Shinomoto, S. 2010. "Kernel bandwidth optimization in spike rate estimation", Journal Comp. Neuroscience, 29, 171-182.



- Shotton, J., Johnson, M., Cipolla, R. 2008. "Semantic texton forests for image categorization and segmentation", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 1-8.
- Shuai, B., Zuo, Z., Wang, B., Wang, G. 2016(a). "DAG-recurrent neural networks for scene labeling", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3620–3629.
- Shuai, B., Zuo, Z., Wang, G., Wang, B. 2016(b). "Scene parsing with integration of parametric and non-parametric models", IEEE Trans. Image Process., 25 (5), 2379-2391.
- Stefanski, J., Mack, B., Waske, B. 2013. "Optimization of object-based image analysis with random forests for land cover mapping", IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens., 6 (6), 2492-2504.
- Tighe, J., Lazechnik, S. 2010. "SuperParsing: scalable nonparametric image parsing with superpixels", Proc. European Conf. Computer Vision (ECCV), 6315, 352-365.
- Tighe, J., Lazechnik, S. 2013(a). "Superparsing: scalable nonparametric image parsing with superpixels", Int. Journal Computer Vision, 101 (2), 329-349.
- Tighe, J., Lazechnik, S. 2013(b). "Finding things: image parsing with regions and per-exemplar detectors", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3001-3008.
- Tighe, J., Niethammer, M., Lazechnik, S. 2015. "Scene parsing with object instance inference using regions and per-exemplar detectors", Int. Journal Computer Vision, 112 (2), 150–171.
- Torralba, A., Fergus, R., Freeman, W.T. 2008. "80 million tiny images: a large dataset for non-parametric object and scene recognition", IEEE Trans. Pattern Anal. Mach. Intell., 30 (11), 1958-1970.
- Vedaldi, A., Soatto, S. 2008. "Quick shift and kernel methods for mode seeking", Proc. European Conf. Computer Vision (ECCV 2008), 5305, 705-718.
- Wang, J., Yang, J., *vd.* 2010. "Locality-constrained linear coding for image classification", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3360-3367.
- Woods, K., Bowyer, K., Kegelmeyer, W.P. 1996. "Combination of multiple classifiers using local accuracy estimates", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 391-396.
- Wu, Z., Hu Z., Fan, Q. 2012. "Superpixel-based unsupervised change detection using multi-dimensional change vector analysis and SVM-based classification", ISPRS Annals Photogrammetry, Remote Sensing Spatial Info. Sciences, I-7, 257-262.
- Yang, J., Price, B., Cohen, S., Yang, M.-H. 2014. "Context driven scene parsing with attention to rare classes", Proc. IEEE Conf. Comp. Vision Pattern Recog. (CVPR), 3294–3301.
- Zheng, S., Jayasumana, S., *vd.* 2015. "Conditional random fields as recurrent neural networks", Proc. IEEE Int. Conf. Computer Vision (ICCV), 1529–1537.
- Zhou, X., Yu, K., Zhang, T., Huang, T.S. 2010. "Image classification using super-vector coding of local image descriptors", Proc. European Conf. Computer Vision (ECCV), 141-154.

**TÜBİTAK**  
**PROJE ÖZET BİLGİ FORMU**

Proje Yürütücüsü:	Prof. Dr. HASAN FEHMİ ATEŞ
Proje No:	115E307
Proje Başlığı:	İmge Ayırıştırma İçin Özgün Süperpiksel Bölütleme Algoritmaları Geliştirilmesi
Proje Türü:	1001 - Araştırma
Proje Süresi:	24
Araştırmacılar:	
Danışmanlar:	
Projenin Yürütüldüğü Kuruluş ve Adresi:	IŞIK Ü.
Projenin Başlangıç ve Bitiş Tarihleri:	01/09/2015 - 01/09/2017
Onaylanan Bütçe:	146887.0
Harcanan Bütçe:	110177.55
Öz:	<p>Süperpikseller görüntü bölütleme ve ayırıştırma problemlerinde yaygın olarak kullanılmaktadır. Sahne etiketlemede görüntü bir süperpiksel algoritması ile görsel olarak tutarlı küçük parçalara bölütlenmekte; daha sonra süperpikseller farklı sınıflara ayırıştırılmaktadır. Bu projede bölütleme ve etiketleme bütünsel bir bakış açısı ile ele alınarak görüntü ayırıştırmanın farklı adımları için özgün yaklaşımlar geliştirilmiştir. Yapılan çalışmalar, süperpikseller için alternatif bölütleme, öznelik çıkarımı, sınıf-olabilirlik hesaplama ve bağlamsal modelleme yöntemleri geliştirilmesini kapsamaktadır. Projede öncelikle farklı bölütleme yöntem ve parametrelerinin etiketleme doğruluğu üzerindeki etkisi test edilmiştir. Daha sonra süperpiksel özneliklerinin seçimi ve kodlanması, sınıf etiketlerinin olabilirlik hesabının modellenmesi üzerinde durulmuştur. Son olarak, alternatif bölütleme sonuçlarının kaynaştırılması için geliştirilmiş bağlamsal modelleme yaklaşımı geliştirilmiştir. Önerilen yöntemler çeşitli anlambilimsel görüntü veritabanlarında test edilmiş ve eniyelenmiştir. Ayrıca projenin son döneminde, yapılan çalışmalar uydu görüntülerinden arazi örtüsü sınıflandırma problemine uyarlanmıştır. Benzetim sonuçları, farklı bölütleme yöntemlerindeki tümler bilginin doğru şekilde birleştirilmesiyle görüntü etiketleme doğruluğunda ciddi artışlar elde edildiğini ortaya koymuştur.</p>
Anahtar Kelimeler:	görüntü ayırıştırma, etiketleme, bölütleme, süperpiksel
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır
Projeden Yapılan Yayınlar:	<p>1- Kernel Likelihood Estimation for Superpixel Image Parsing (Bildiri - Uluslararası Bildiri - Poster Sunum), 2- Kernel Likelihood Estimation for Superpixel Image Parsing (Bildiri - Uluslararası Bildiri - Poster Sunum), 3- Kernel Kod-tablosu Kodlaması ile Sahne Etiketleme (Bildiri - Ulusal Bildiri - Sözlü Sunum), 4- Improving Semantic Segmentation with Generalized Models of Local Context (Bildiri - Uluslararası Bildiri - Poster Sunum),</p>