

Multispektral Görüntülerde İnce Taneli Nesne Tanıma ve Örneksiz Öğrenme

Fine-Grained Object Recognition and Zero-Shot Learning in Multispectral Imagery

Gencer Sümbül ve Selim Aksoy
Bilgisayar Mühendisliği Bölümü
Bilkent Üniversitesi, Ankara, Türkiye
gencer.sumbul@bilkent.edu.tr, saksoy@cs.bilkent.edu.tr

Ramazan Gökberk Cinbiş
Bilgisayar Mühendisliği Bölümü
Orta Doğu Teknik Üniversitesi, Ankara, Türkiye
gcinbis@ceng.metu.edu.tr

Özetçe —Multispektral görüntülerde bir nesne sınıfı için çok sayıda alt kategori içinde o nesnenin türünün tanımlanmasını hedefleyen ince taneli nesne tanıma problemi ve örneksiz öğrenme senaryosu için bir yöntem sunulmaktadır. Örneği bulunan sınıflar ile yeni ve görülmemiş sınıflar arasında bir ilişki kurabilmek için evrimsel sinir ağından çıkarılan görüntü öznetelikleri ile sınıfların yardımcı bilgileri arasında bir uyumluluk fonksiyonu öğrenilmektedir. Örneği bulunmayan sınıflar için bilgi aktarımı, bu fonksiyonun en büyüklüğü ile gerçekleştirilmektedir. Yardımcı bilgi olarak elle etiketlenmiş nitelikler, doğal bir dil modeli ve bilimsel taksonomi ile değerlendirilen modelin performansı (%15.2) 16 test sınıfı için diğer yöntemlerden daha iyidir.

Anahtar Kelimeler—Örneksiz öğrenme, ince taneli sınıflandırma, nesne tanıma

Abstract—We present a method for fine-grained object recognition problem, that aims to recognize the type of an object among a large number of sub-categories, and zero-shot learning scenario on multispectral images. In order to establish a relation between seen classes and new unseen classes, a compatibility function between image features extracted from a convolutional neural network and auxiliary information of classes is learnt. Knowledge transfer for unseen classes is carried out by maximizing this function. Performance of the model (15.2%) evaluated with manually annotated attributes, a natural language model, and a scientific taxonomy as auxiliary information is promisingly better than the other methods for 16 test classes.

Keywords—Zero-shot learning, fine-grained classification, object recognition

I. GİRİŞ

Uzaktan algılanmış görüntülerin uzamsal ve spektral çözünürlüğünün artması ile birlikte havadan ve uydudan çekilmiş görüntülerde yeni ayrıntılar ortaya çıkmış, bu sayede yeni nesne sınıflarının algılanması ve tanınması kolaylaşmıştır. Böylece, uzaktan algılanmış görüntü analizinde en popüler problemlerden biri otomatik nesne tanıma olmuştur. Hem geleneksel öznetelikleri kullanan sınıflandırıcılar, hem de sınıflandırma kurallarıyla birlikte öznetelikleri de öğrenmeyi amaçlayan derin sinir ağları, farklı kaynaklardan gelen veri

kümelerinde dikkat çekici bir performans sergilemiştir. Bu tür verilerin ortak özelliği, güdümlü bir öğrenme hedefine yetecek kadar verinin genellikle bulunduğu, kentsel, kırsal ve tarımsal alanlara ait görece ayrıntılı sınıfları içermeleridir. Bu tür veri kümeleri, farklı yaklaşımların nesnel olarak karşılaştırılmasını sağlayarak başarı seviyelerini ilerletmede faydalı olmuştur. Ancak, uzaktan algılanmış görüntülerin çeşitliliği halen birçok araştırma problemine yol açmaktadır.

Bir nesne sınıfı için çok sayıda alt kategori içinde o nesnenin türünün tanımlanmasıyla ilgilenen *ince taneli nesne tanıma* problemi uzaktan algılama literatüründe çok az ilgi görmüş konulardan bir tanesidir. Bu tip problemlerde alt kategorilerin ayrıştırılması çok zor olabilmektedir. Bununla birlikte, varolan sınıflandırma modellerinin bu tür bir tanıma görevine nasıl yaklaşacağı belirsizliğini korumaktadır. İnce taneli nesne tanıma görevi, diğer sınıflandırma ve tanıma görevlerinden iki önemli açıdan farklıdır: küçük örneklem büyüklüğü ve sınıf dengesizliği. Uzaktan algılama geleneksel olarak bol miktarda veriden yararlanmasına rağmen, etiket bilgisini elde etmek sınıflandırma çalışmalarında daima önemli bir tikanıklık yaratmıştır. Sınırlı eğitim verisinin etkilerini azaltmaya yönelik genel girişimler, parametre kestiriminde ve öznetelik bulmada düzleştirme [1], sınıflandırıcı topluluklarını kullanmayı ve uzamsal ve bağlamsal bilgilerden faydalanmayı [2] içermektedir. Ancak, ince taneli tanıma görevlerindeki düşük sınıflar arası değişim ve yüksek sınıf içi değişim istatistiksel çözümlerin kullanımını sınırlamaktadır. Etiketli örnek yetersizliğinin üstesinden gelmek için bir diğer yaklaşım, etkileşimli yeni örnekler toplamak için aktif öğrenmeyi kullanmaktır [3]. Bununla birlikte, ince taneli nesne tanıma, verileri görsel olarak inceleyerek çok sayıda ve çok benzer nesne kategorileri için örnek toplama alan uzmanları için bile çok zor olabilmektedir. Dahası, uzamsal olarak dağıtık verileri edinme maliyetleri, saha ziyaretleri yoluyla örnek toplama işlemini pratik olarak olanaksız kılabilir. Ayrıca, eğitim verilerindeki sınıf dengesizliği de güdümlü öğrenme sırasında sorunlara neden olabilmektedir.

Bu sorunların yanı sıra, daha da zor bir senaryo olarak sınıfların bazıları için *hiçbir eğitim örneği bulunmayan örneksiz öğrenme* (zero-shot learning, ZSL) problemi ele alınabilir. Örneksiz öğrenme, daha önce öğrendiği kategorilerle, eğitim

örnekleri olmayan yeni kategorileri ilişkilendirerek bir tanıma modeli oluşturmayı amaçlamaktadır [4]. Örneksiz öğrenmede test kategorileri için herhangi bir eğitim örneği mevcut olmadığından, tek başına görüntü verileri *eğitimde görülmeyen ve eğitimde görülen* sınıflar arasındaki ilişkiyi oluşturmak için yeterli değildir. Bu nedenle, bu ilişkiyi kurmak için ara katman olarak işlev görebilecek yeni yardımcı bilgi kaynakları kullanılmalıdır.

Uzaktan algılama literatüründe, ince taneli nesne tanıma için örneksiz öğrenme problemi [5]'te RGB görüntüleri üzerinde incelenmiştir. Bu bildiride, örneksiz öğrenme ile ince taneli nesne tanıma problemi uzaktan algılanmış multispektral görüntülerde incelenmektedir. Önerilen yaklaşım, girdi görüntü verilerinde gözlemlenen görsel özellikler ile sınıfların anlamsal özelliklerini tanımlayan yardımcı bilgiler arasındaki uyumluluğu modelleyen çift doğrusal bir fonksiyon kullanmaktadır. Görüntü içeriği, eğitim verisinde görülen sınıflardan öğrenilen bir evrişimsel sinir ağı kullanılarak çıkarılmış öznitelikler tarafından modellenmiştir. Kullanılan yardımcı bilgi üç tamamlayıcı alandan toplanmıştır. Bunlar, alan uzmanlığını yansıtan elle etiketlenmiş nitelikler, büyük metin derlemesi üzerinde eğitilmiş doğal bir dil modeli ve bilimsel taksonominin hiyerarşik gösterimidir. Öğrenme aşamasında en büyük olabilirlik ilkesini kullanarak eğitimde görülen sınıflardan uyumluluk fonksiyonunun nasıl kestireleceği ve bu fonksiyonun en yüklenmesi ile bilgi aktarımının nasıl gerçekleştirileceği gösterilmektedir. Ardından, önerilen yöntem farklı ağaç türlerini içeren bir veri kümesinde denenmiştir.

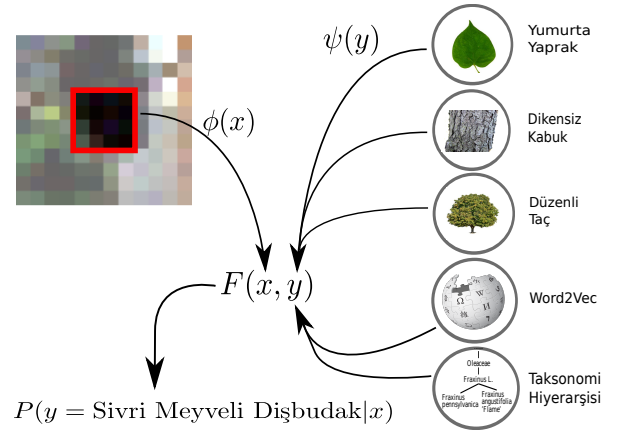
Önerilen yöntemin ayrıntıları Bölüm II'de açıklanmıştır. Deneyisel çalışmalar Bölüm III'te verilmiştir. Bölüm IV'te ise sonuçlar sunulmuştur.

II. YÖNTEM

A. Örneksiz Öğrenme Modeli

Amacımız verilen bir görüntüyü, $x \in \mathcal{X}$, hedef sınıflardan birine, $y \in \mathcal{Y}$, eşleyen ayırtıcı bir fonksiyon öğrenmektir. Burada $\mathcal{X}$ tüm görüntülerin uzayı ve $\mathcal{Y}$ tüm nesne sınıflarının kümesidir. Örneksiz öğrenmenin tanımına göre, eğitim örnekleri sadece eğitimde *görülen sınıflardan* alınabilir. Bu sınıflar, bütün sınıfların bir alt kümesidir ($\mathcal{Y}_{tr} \subset \mathcal{Y}$). Bu nedenle, hiçbir eğitim örneği olmayan *görülmemiş sınıfları* ($\mathcal{Y}_{te} \subset \mathcal{Y}$ ve $\mathcal{Y}_{tr} \cap \mathcal{Y}_{te} = \emptyset$) tanıyabilen bir model oluşturmamız geleneksel güdümlü yöntemlerle doğrudan mümkün değildir. Bu zorluğun üstesinden gelmek için öncelikle *sınıf gösterimi* olarak adlandırılan bir vektör uzayı temsiline her sınıf için mevcut olduğu varsayılmaktadır. Her bir sınıf gösterimi vektörünün sınıfın (görsel) özelliklerini betimlemesi beklenir ve böylece sınıflandırma bilgisi, görülen sınıflardan görülmeyen sınıflara aktarılabilir. Bu bilgi aktarımını gerçekleştirmek için bir $F : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ uyumluluk fonksiyonu kullanılmaktadır. Bu fonksiyon, verilen görüntü-sınıf ikilisini (x, y) sayıl bir değere eşlemektedir. Bu değer, x görüntüsünün y sınıfına atanmasına olan güveni temsil eder. Eğitimde görülmemiş sınıfların örneklerini tanımak için kullanılacak olan uyumluluk fonksiyonunu öğrenirken sadece görülen sınıfların örneklerini kullanabildiğimiz için, $F(x, y)$ fonksiyonu sınıf-agnostik bir model kullanılmalıdır. [6] çalışmasını takiben uyumluluk fonksiyonu çift doğrusal yapıda aşağıdaki gibidir:

$$F(x, y) = \phi(x)^T W \psi(y). \quad (1)$$



Şekil 1: Önerdiğimiz yöntem, $\phi(x)$ görüntü ve $\psi(y)$ sınıf gösterimi arasındaki uyumluluk fonksiyonunu, $F(x, y)$, öğrenmektedir. $F(x, y)$, eğitimde görülmeyen sınıfların örneklerini tanımda kullanılmaktadır.

Bu denklemde $\phi(x)$, görüntü gösterimi olarak adlandırılan d boyutlu bir görüntü temsiline, $\psi(y)$, m boyutlu bir sınıf gösterimini, ve W , $d \times m$ boyutlu bir matrisi ifade etmektedir. Bu uyumluluk fonksiyonu, görüntü ve sınıf gösterimleri arasındaki çapraz-alan ilişkisinin sınıf-agnostik bir modeli olarak düşünülebilir. Yöntemin görselleştirmesi için Şekil 1'e bakılabilir.

Uyumluluk fonksiyonunu öğrenmek için, çok sınıflı lojistik regresyon sınıflandırıcısının uyarlamasına dayanan bir yöntem geliştirilmiştir. Sınıfa ait sonsal olasılık dağılımı, uyumluluk değerlerinin *softmax* sonucu olarak tanımlanmaktadır:

$$p(y|x) = \frac{\exp(F(x, y))}{\sum_{y' \in \mathcal{Y}_{tr}} \exp(F(x, y'))}. \quad (2)$$

Bu denklemde $\mathcal{Y}_{tr} \subset \mathcal{Y}$, eğitimde görülen sınıfların kümesini ifade etmektedir. Ardından, verilen N_{tr} eğitim örneği için, $F(x, y)$ uyumluluk fonksiyonunun en büyük olabilirlik ilkesi ile öğrenilmesi amaçlanmaktadır. Veri kümesinin bağımsız özdeşçe dağılmış örneklerden oluştuğunu varsayarsak, etiket olabilirliği şu şekilde bulunabilir:

$$\text{enbüyütme}_{W \in \mathbb{R}^{d \times m}} \prod_{i=1}^{N_{tr}} p(y_i | x_i). \quad (3)$$

En iyileme problemi, eğitim örneklerinin öngörülen gerçek sınıf olasılıklarını en büyükleyen W matrisinin bulunması olarak yorumlanabilir. Eşdeğer olarak, negatif logaritmik olabilirliği en küçültmek parametreler bulunabilir:

$$\text{enküçültme}_{W \in \mathbb{R}^{d \times m}} \sum_{i=1}^{N_{tr}} -\log p(y_i | x_i). \quad (4)$$

Yerel bir en iyi çözüm bulmak için, stokastik gradyan inişi (SGİ) tabanlı en iyileme kullanılmaktadır. Bu yöntemin ana fikri, bir dizi eğitim örneğini yinelemeli örneklemek, yığın üzerinde yaklaşık gradyanı hesaplamak ve yaklaşık gradyanı kullanarak model parametrelerini güncellemektir. Bizim durumumuzda, SGİ'nin t yinelemesinde, eğitim örneklerinin bir B_t yığını üzerindeki G_t gradyan matrisi aşağıdaki gibi hesaplanabilir:

$$G_t = - \sum_{i \in B_t} \nabla_W \log p(y_i | x_i).$$

Bu denklemde, i örneği için logaritmik olabilirlik teriminin gradyanı şöyledir:

$$\nabla_W \log p(y_i|x_i) = \phi(x_i)\psi(y_i)^\top - \sum_{y \in \mathcal{Y}_r} p(y|x_i)\phi(x_i)\psi(y)^\top.$$

Verilen yaklaşık gradyan için sade SGİ algoritması, G_t 'ye orantılı bir matrisi model parametrelerinden çıkararak çalışmaktadır:

$$W_t \leftarrow W_{t-1} - \alpha G_t. \quad (5)$$

Bu denklemde W_t , güncellenmiş model parametrelerini ve α (öğrenme hızı), SGİ yinelemeleri boyunca güncelleme oranını belirtmektedir.

B. Görüntü gösterimi

Uzaktan algılanmış multispektral görüntülerde görsel içerik temsillerini öğrenmek ve çıkarmak için derin bir evrişimsel sinir ağı (convolutional neural network, CNN) kullanılmaktadır. CNN kullanılmasının nedeni piksel tabanlı spektral bilgileri ve uzamsal doku içeriğini modelleyebilmektir. Kullanılan mimariye sırasıyla 5×5 , 5×5 ve 3×3 boyutlu filtreler içeren üç evrişimsel katman ve tümü birbirine bağlı sinirlerden oluşan 2 katman bulunmaktadır. Uzamsal çözünürlüğü kaybetmemek adına CNN modellerinde yaygın olarak kullanılan biriktirme (pooling) katmanı bu çalışmada tercih edilmemiştir. Bir dizi benzer derin ve geniş mimari incelenmiş, sunulan mimariye ek olarak 3×3 , 3×3 ve 3×3 boyutlu filtrelerin bulunduğu evrişimsel katmanlara sahip mimarinin sonuçları da paylaşılmıştır. CNN modeli, eğitimde görülen sınıflar üzerinde ve logaritmik etiket olabilirliğini en büyütecek şekilde eğitilmiştir. Sinir ağı eğitildikten sonra, örneksiz öğrenme modelindeki görüntü gösterimi, $\phi(x)$, için ilk tam bağlantılı katmanın çıktısı olan 128 boyutlu vektör kullanılmıştır.

C. Sınıf gösterimi

Sınıf gösterimi, sınıflandırma bilgisinin aktarılması için bilgi kaynağı olduğundan sınıfların görsel özelliklerini ifade etmelidir. Bu amaçla üç farklı sınıf gösterimi yönteminin birleşimi kullanılmaktadır: (i) hedef tema alanından topladığımız elle etiketlenmiş nitelikler, (ii) güdümsüz dil modelleri kullanılarak oluşturulmuş metin gösterimleri ve (iii) bilimsel bir taksonomiye dayalı hiyerarşik gösterim. Görsel nitelikler, nesnelerin parça, doku ve şekil gibi görsel olarak ayırıcı özelliklerini belirleyerek elde edilmektedir. Bu çalışmada nitelikleri kullanmak için, Florida Ağaçları bilgi metnine [7] dayanarak, ağaç türleri için 25 nitelik (yükseklik, taç yoğunluğu, yaprak şekli, çiçek rengi gibi) toplanmıştır. Nitelikleri tamamlayan ikinci yöntem olarak, kelimeler ve onların bağlamları arasındaki ilişkiyi modelleyen güdümsüz Word2Vec yaklaşımı [8] kullanılmaktadır. Word2Vec tarafından kodlanan anlamsal ilişkileri kullanmak için, Wikipedia makaleleri üzerinde eğitilmiş 1000 boyutlu bir gösterim modeli kullanılmakta ve ağaç türlerinin yaygın adlarının kelime gösterimleri çıkarılmaktadır. Birden fazla kelimeye sahip kategorilerde, her kelime için çıkartılan gösterim vektörlerinin ortalaması alınmaktadır. Kullanılan üçüncü ve son sınıf gösterimi türü, ağaç türleri arasındaki benzerlikleri bilimsel sınıflandırmaya dayanarak ifade etmeyi amaçlamaktadır. Taksonomi ağacındaki ağaç türlerinin konumunu ve atalarını ifade etmek için, [9]'de açıklanan ağaç-vektör dönüşüm şeması uygulanmaktadır. Verilen bir ağaç

türüne karşılık gelen gösterim vektörü, her bir düğüm için ikili bir değer tanımlayarak ve yalnızca kökten ilgili yaprak düğümüne giden yolda görünen düğümlere karşılık gelen değerleri 1 yaparak elde edilmektedir. Bu üç gösterim yöntemiyle üretilen vektörler art arda eklenerek son gösterim vektörü oluşturulmaktadır. Sınıf gösterim yöntemlerinin detayları [5]'te verilmiştir.

III. DENEYLER

A. Deneysel düzeneği

Veri kümesi oluşturulurken, [5]'te kullanılan, 40 sınıfa ait sokak ağaçlarının konum bilgisi ve isimlerinin bulunduğu GIS verisi kullanılmıştır. Görüntü verisi için aynı çalışmada kullanılan RGB görüntüsü yerine, aynı uzamsal alanı içeren WorldView-2 uydusundan çekilmiş 8 bantlı ve 1.84 metre uzamsal çözünürlüğe sahip multispektral görüntü kullanılmıştır. Dolayısıyla her ağaç, 25×25 boyutlu pencereler yerine aynı alana denk gelecek şekilde 4×4 boyutlu pencerelerle temsil edilmiştir.

Deneysel düzeneği güdümlü ve örneksiz öğrenmeyi birlikte içermektedir. Dolayısıyla, önyargısız değerlendirme sonuçları elde etmek için veri kümemizdeki 40 sınıf, ortak sınıf olmayacak şekilde ayrı üç kümeye bölünmüştür. *Güdümlü kümeye* 18 sınıf, *ZSL sağlama kümesine* 6 sınıf ve geri kalan *ZSL test kümesine* 16 sınıf ayrılmıştır. Bölmeler kabaca her sınıftaki örnek sayısına göre ayarlanmıştır. ZSL doğruluk değerlendirmesinde güvenilir bir performans elde edebilmek için, çoğunlukla en büyük sınıflar güdümlü kümeye, en küçük sınıflar ZSL sağlama kümesine, geri kalanlar ise ZSL test kümesine ayrılmıştır. Güdümlü küme iki amaç için kullanılmaktadır: (i) CNN modelini güdümlü sınıflandırmada değerlendirmek ve (ii) güdümlü kümeye ait sınıfları kullanarak ZSL modelini eğitmek. Güdümlü sınıflandırma deneyleri için sadece güdümlü küme sınıfları kullanılmaktadır. Bu sınıflara ait görüntüler, *güdümlü eğitim* (60%), *güdümlü sağlama* (20%) ve *güdümlü test* (20%) olarak alt kümelere bölünmektedir. Bu üç alt küme, 18 güdümlü küme sınıfına ait görüntüleri içermektedir ve ZSL sağlama kümesinden veya ZSL test kümesinden bir sınıfa ait herhangi bir görüntü içermemektedir. Güdümlü sınıflandırıcıların parametreleri seçilirken, güdümlü sağlama kümesindeki performans en iyileştirilmeye çalışılmıştır. ZSL deneylerinde, güdümlü kümedeki tüm görüntüler kullanılarak ZSL modeli eğitilmiştir. ZSL modelinin parametrelerini ayarlamak için ZSL sağlama kümesinde örneksiz tanıma doğruluğu kullanılmıştır. Model performansı son olarak eğitimde görülmeyen sınıfları içeren ZSL test kümesinde değerlendirilmiştir. Deneyler boyunca performans ölçütü olarak, her sınıfın doğruluk oranının ortalaması alınarak bulunan düzgelemli doğruluk kullanılmıştır. Bu şekilde, çok sayıda örneğe sahip sınıflara yönelik önyargılardan kaçınmak hedeflenmiştir.

B. Güdümlü ince taneli sınıflandırma

ZSL sonuçlarını sunmadan önce, CNN modeli rastgele orman sınıflandırıcısı ile güdümlü bir şekilde Tablo I'de karşılaştırılmıştır. Adil bir karşılaştırma için, her iki yöntem de güdümlü eğitim kümesi üzerinde eğitilmiş ve parametreleri güdümlü sağlama kümesinde ayarlanmıştır. Ayrıca her bir eğitim örneği görüntüsü, yükseklik ve genişlik %0 – %20 aralığında rastgele kaydırılarak değiştirilmiştir. Sonuçlardan,

TABLE I: NESNE
TANIMA SONUÇLARI (%)

<i>Güdümlü Sınıflandırma</i>	
Rastgele Tahmin	5.6
Rastgele Orman (RGB) [5]	15.8
Rastgele Orman (Multispektral)	21.3
CNN (RGB) [5]	34.6
CNN (Multispektral) Filtre: $(5 \times 5), (5 \times 5), (3 \times 3)$	36.7
CNN (Multispektral) Filtre: $(3 \times 3), (3 \times 3), (3 \times 3)$	39.0
<i>Örneksiz Öğrenme (ZSL)</i>	
Rastgele Tahmin	6.3
ZSL (RGB) [5]	14.3
ZSL (Multispektral) Filtre: $(5 \times 5), (5 \times 5), (3 \times 3)$	14.7
ZSL (Multispektral) Filtre: $(3 \times 3), (3 \times 3), (3 \times 3)$	15.2

sınıflandırma yöntemlerinin rastgele tahmin temelinden (%5.6) daha iyi performans gösterdiği görülmektedir. Ayrıca, RGB (%34.6) ve multispektral (%36.7 ve %39.0) görüntüye uygulanmış CNN modelleri, rastgele ormandan (RGB için %15.8, multispektral için %21.3) daha iyi çalışmaktadır. Hem rastgele orman sınıflandırıcısı hem de farklı filtre boyutlarında CNN için, multispektral görüntünün RGB görüntüsünden daha iyi sonuç verdiği gözlemlenmektedir. Bu sonuçlar, CNN yaklaşımıyla öğrenilmiş derin görüntü temsiliinin avantajını ve uzamsal çözünürlüğü RGB görüntüye göre düşük olmasına rağmen multispektral verideki spektral çözünürlüğün önemini göstermektedir. Buna ek olarak, binaları ağaçlardan ya da yolları çimden vb. ayırmaya çalışan geleneksel sınıflandırma senaryolarından oldukça farklı olan ince taneli sınıflandırmanın ne kadar zor bir problem olduğu gözlemlenebilmektedir.

C. İnce taneli örneksiz öğrenme

ZSL modeli, güdümlü küme üzerinde eğitilmiş ve tüm model parametreleri ZSL sağlama kümesindeki düzgelenmiş doğruluğa göre ayarlanmıştır. Ayrıca eğitim sınıflarındaki örnek sayısı dengesizliğinin ZSL modelini olumsuz etkilediği gözlemlenmiştir. Bu sorunu hafifletmek için, her sınıfın eğitim kümesi büyüklüğü en büyük sınıfın büyüklüğüne eşit olacak şekilde rastgele ek örnekleme uygulanmıştır. 16 ZSL test sınıfı üzerindeki ZSL sonuçları Tablo I'de sunulmuştur. ZSL modelinin rastgele tahmin temelinden (%6.3) daha iyi performans gösterdiği görülmektedir. Dolayısıyla, olasılıksal ZSL formüleştirmesinin geçerliliği doğrulanmaktadır.

Görüntü gösteriminin, ZSL performansı üzerinde derin bir etkisi vardır. RGB görüntüsü için kullanılan mimariyle elde edilen multispektral görüntü gösteriminin ZSL sonucu (%14.7) ile yaklaşık 6 kat daha düşük uzamsal çözünürlüğe rağmen aynı başarı yakalanmıştır. Bu, spektral bilginin görüntü gösterimi için önemini belirtmektedir. Multispektral görüntüye daha uygun filtre boyutu (3×3) tercih edildiği zaman, başarının arttığı (%15.2) gözlemlenmiştir. Bu sonuç, veriye ve alana özgü CNN mimarisi seçmenin ZSL başarısı için önemini ortaya koymaktadır.

TABLE II: FARKLI
SINIF GÖSTERİMLERİ
İÇİN ÖRNEKSİZ
ÖĞRENME
SONUÇLARI (%)

Nitelik	Hiyerarşi	Word2Vec	
✓	×	×	13.3
×	✓	×	13.0
×	×	✓	13.3
✓	✓	×	13.5
✓	×	✓	13.4
×	✓	✓	14.6
✓	✓	✓	15.2

Sınıf gösterimi üç farklı gösterim yönteminin birleşimi olarak kullanılmıştır. Her birinin katkısını anlamak için, Tablo II'de sınıf gösterimi yöntemlerinin farklı birleşiminin ZSL performansı sunulmuştur. Burada (3×3) CNN filtresi kullanılmıştır. Tablonun ilk üç satırında, gösterim yöntemleri tek tek kullanıldığında nitelik (%13.3), hiyerarşi (%13.0) ve Word2Vec (%13.3) performanslarının benzer olduğu görülmektedir. Ardından gelen üç sıra, hiyerarşi-Word2Vec (%14.6) gösterim çiftinin, tek tek gösterimlerle birlikte nitelik-hiyerarşi (%13.5) veya nitelik-Word2Vec (%13.4) çiftlerine kıyasla daha iyi sonuç verdiğini göstermektedir. Tüm gösterimlerin birleşimine dayanan ve en yüksek doğruluğu veren sonuç (%15.2) sınıf gösterimlerinin birbirlerini tamamlayıcı nitelikte olduğunu göstermektedir.

IV. SONUÇ

Bu çalışmada, multispektral görüntülerde ince taneli nesne tanıma için örneksiz öğrenme problemi üzerinde çalışılmıştır. Eğitim örnekleri bulunmuyor olsa bile çok benzer sınıfları tanıma problemi ile başa çıkabilmek için yardımcı bilgi kaynaklarından yararlanılmıştır. Önerilen yaklaşım ile görülen sınıflardan çift doğrusal bir fonksiyon öğrenilmiştir. Böylece, girdi görüntü verilerinde gözlemlenen görsel özellikler ile sınıfların anlamsal özelliklerini tanımlayan yardımcı bilgiler arasındaki uyumluluk modellenmiştir. Ardından uyumluluk fonksiyonunun bilgi aktarımını gerçekleştirmek için nasıl kullanılabileceği gösterilmiştir. Gözlemlenen deney sonuçları ile, 6 kat daha düşük uzamsal çözünürlüğe sahip multispektral verinin, uygun bir CNN mimarisi ile RGB verisinden daha iyi örneksiz öğrenme başarısı sağlayabileceği gösterilmiştir.

TEŞEKKÜR

Bu çalışma TÜBİTAK 116E445 numaralı proje ve Bilim Akademisi BAGEP Ödülü ile desteklenmiştir.

KAYNAKLAR

- [1] B.-C. Kuo and K.-Y. Chang, "Feature extractions for small sample size classification problem," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 756–764, March 2007.
- [2] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, "Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, 2015.
- [3] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, March 2011.
- [4] B. Romera-Paredes and P. H. S. Torr, "An embarrassingly simple approach to zero-shot learning," in *Intl. Conf. Mach. Learn.*, vol. 37, 2015, pp. 2152–2161.
- [5] G. Sumbul, R. G. Cinbis, and S. Aksoy, "Fine-grained object recognition and zero-shot learning in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 770–779, February 2018.
- [6] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning — the good, the bad and the ugly," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017.
- [7] University of Florida. (2016, October) 680 tree fact sheets: Trees by common name. [Online]. Available: http://hort.ifas.ufl.edu/database/trees/trees_common.shtml
- [8] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
- [9] R. Mittelman, M. Sun, B. Kuipers, and S. Savarese, "A Bayesian generative model for learning semantic hierarchies," *Frontiers in Psychology*, vol. 5, p. 417, May 2014.