

İmge Sınıflandırması için Yeni Öznitelik Çıkarım Yöntemi: ADD-TDA Algısal Özet Fonksiyonu Tabanlı Evrişimsel Sinir Ağ (ADD-TDA-ESA)

A New Feature Extraction Method for Image Classification: DWT-SVD Perceptual Hash-Based Convolutional Neural Network (DWT-SVD-CNN)

Fatih ÖZYURT
Fırat Üniversitesi
fatihozyurt@firat.edu.tr

Engin AVCI
Fırat Üniversitesi
enginavci@firat.edu.tr

Öz

Bu çalışma, Evrişimsel Sinir Ağı (ESA) kullanarak, imge sınıflandırma süresini azaltan, sınıflandırma performansını kabul edilebilir değerde tutabilen bir yöntemi önermektedir. Ayrık Dalgacık Dönüşümü - Tekil Değer Ayırıştırma'ya dayalı algısal özet fonksiyonu kullanarak Evrişimsel Sinir Ağı (ADD-TDA-ESA) adlı melez modelde, sınıflandırma süresini azaltmak için ESA ile birlikte ADD-TDA tabanlı algısal özet fonksiyonu kullanılmıştır. Algısal özet fonksiyonlarının en önemli özelliği imgelerin belirgin özelliklerini elde etmektir. Bu yöntemde, ilk olarak imgelerin belirgin özelliklerini elde etmek için ADD-TDA algısal özet fonksiyonu uygulanmıştır. Daha sonra belirgin özelliklerden oluşan 32x32 boyutundaki imgeler ESA'ya girdi olarak verilerek öznitelikler çıkartılıp Destek Vektör Makinesine sınıflandırma için verilmiştir. ADD-TDA-ESA yöntemi, Caltech-101 veri tabanında bulunan imgeler için uygulanmıştır. Deneysel sonuçlar önerilen ADD-TDA-ESA yönteminin %95.8 yüksek doğruluk oranına sahip olduğunu göstermiştir. Ayrıca kullanılan bu yöntem ile klasik yöntemde 241.21 saniye olan çalışma süresi 83.08 saniyeye düşmüştür. Deneysel sonuçları ADD-TDA-ESA yönteminin, imge sınıflandırma doğruluğunu yüksek tutarak klasik ESA'ya göre çok daha hızlı performans sergilediğini göstermektedir.

Anahtar Sözcükler: Derin Öğrenme, Evrişimsel Sinir Ağları, Sınıflandırma, Algısal Özet Fonksiyonu, Ayrık Dalgacık Dönüşümü, Tekil Değer Ayırıştırma

Abstract

This paper proposes a method, which reduces the image classification time and maintains the classification performance above an acceptable threshold by using Convolutional Neural Network (CNN). A hybrid model called Discrete Wavelet Transform - Singular Value Decomposition based Perceptual Hash Convolutional Neural Network (DWT-SVD-PH-CNN) is proposed by using a perceptual hash function together with CNN to reduce the classification time. The most important feature of perceptual hash functions is to obtain the salient features of images. First, DWT-SVD based perceptual hash function is applied to images for obtaining salient features. Then, images making up of salient features, are produced in 32x32 format and given as inputs to CNN, where Support Vector Machine (SVM) is used to classify the images. In this paper, the DWT-SVD-PH-CNN method is applied to Caltech 101 image database. Experimental results show that the proposed DWT-SVD-PH-CNN method has a high accuracy, about 95.8 %. Moreover, this method reduces the execution time from 241.21 seconds to 83.08 seconds compared to the classical method. Thus, the experimental results show that the proposed DWT-SVD-PH-CNN method performs much faster than classical CNN by maintaining the image classification accuracy high.

Gönderme ve kabul tarihi: 26.08.2018-04.12.2018
F. Özyurt: orcid.org//0000-0002-8154-6691
E. Avci: orcid.org/0000-0001-6278-3221
Makale türü: Araştırma

Key words: Deep Learning, Convolutional Neural Network, Classification, Perceptual Hash Function, Discrete Wavelet Transform, Singular Value Decomposition

1. Giriş

Son zamanlarda, imge sınıflandırma ve örüntü tanıma yoğun olarak çalışılan bir alan haline gelmiştir. Klasik olarak özelliklerin çıkarılması için uygulanan en yaygın yöntemler Ölçek değişmeyen özellik dönüşümler (SIFT) [1], Gabor özellikler [2] ve yerel ikili modeli (LBP)'dir [3]. Bu yöntemler, doku sınıflandırma, yüz tanıma ve nesne tanıma gibi farklı sınıflandırma görevlerinde iyi sonuçlar elde edilebilmektedir. Bu yöntemlerin yetersiz kaldığı durumlarda farklı öznelik çıkarımları için uzmanlığa ihtiyaç duyulmaktadır. Klasik olarak çıkartılan özneliklere göre sınıflandırılma yapılmış birçok makale bulunmaktadır [1-6].

Derin öğrenme algoritmaları, imge özelliklerinin klasik olarak çıkartılmasının üstesinden gelmek için yaygın olarak kullanılmaktadır [7-12]. Derin sinir ağları, alt düzey özelliklerden üst düzey özellikler oluşturarak öğrenebilen makine sınıflandırmalarıdır [13]. Derin öğrenme mimarileri, Evrişimsel Sinir Ağı, Seyrek Oto-Kodlayıcılar [14], Gürbüz Boltzmann Makineleri [15,16] ve Derin İnanç Ağı [17] gibi kategorilere ayrılmıştır.

Evrişimsel Sinir Ağı, işlem katmanları (doğrusal ve doğrusal olmayan) tarafından öğrenilebilen çeşitli algoritmalarından oluşmaktadır [18]. Özneliklerin oluşumu bu katmanlar arasındadır[19]. ESA, imge tanıma, segmentasyon ve sınıflandırma alanlarında ciddi performansa sahip olduğu için diğer klasik yöntemlere göre üstün performans gösterir.

Bu makalenin amacı, 7 farklı sınıftan oluşan toplam 3150 imgenin sınıflandırmasını yaparken, algısal özet fonksiyonları kullanarak test doğruluk oranını kabul edilebilir bir değerde tutup çalışma süresini azaltmaktır. Bu nedenle doğru tanıma ve yorumlama performansını azaltmayan fakat çalışma süresini azaltan çok boyutlu renkli imgeleri indirgeyerek belirgin özellikleri elde etmemize sağlayan algısal özet fonksiyonu tabanlı evrişimsel sinir ağı önerilmiştir. İmge yapısını bozduğu için algısal olmayan (kriptolojik) özet fonksiyonları kullanılmamıştır. Önerdiğimiz yöntem ile derin öğrenme mimarilerinden olan ESA ilk defa algısal özet fonksiyonları ile birlikte kullanılmıştır.

Bu çalışmanın katkıları aşağıdaki gibi sıralanabilir:

1 -ADD-TDA tabanlı algısal özet fonksiyonu kullanılarak görüntülerin belirgin özelliklerinden elde edilen 32x32 boyutundaki imgeler evrişimsel sınıflandırmaya giriş olarak verilip öznelikler çıkartıldıktan sonra sınıflandırma işlemi yapılmaktadır. ADD, JPEG sıkıştırmasına, TDA ise döndürmeye karşı dayanıklı özellikler elde etmek için kullanılır.

2 -Derin öğrenmede başarılı olan evrişimsel sinir ağı yapısı kullanılmıştır. Bu yapı, imgelerde öznelik çıkarma safhasında çok başarılıdır. Ancak imgelerde piksel seviyesinde işlem yapması nedeniyle büyük verilerle kullanıldığında çalışma zamanı kısıdına yol açmaktadır. Bundan dolayı, 400x600x3 ile 280x186x3 boyutları arasında değişen toplam 3150 renkli imge 32x32 boyutundaki gri imgelere çevrilerek sınıflandırma işlemi için çalışma süresi ciddi oranda azalmıştır.

3-Bu çalışmada, algısal özet fonksiyonu ve evrişimsel katmanlar kullanılarak öznelik çıkaran hibrid bir yöntem önerilmiştir. Önerilen yöntemde destek vektör makinesi sınıflandırıcı olarak kullanılmıştır.

Makalenin 2. bölümü, ADD-TDA tabanlı algısal özet fonksiyonunu açıklamaktadır. Temel ESA tanımı ve Alexnet modeli 3. bölümde verilmiştir. Önerilen ADD-TDA-ESA yöntemi bölüm 4'te açıklanmıştır. 5. Bölümde deneysel sonuçlar sunulmuştur. Son olarak 6. bölümde ise elde edilen bu sonuçlar yorumlanmıştır.

2. Geçmiş Çalışmalar

Evrişimsel sinir ağları kullanılarak gerçekleştirilen imge işleme çalışmalarında, en önemli problem; yapılan işlemlerin pikseller mertebesinde olmasından dolayı hesaplama maliyetinin yüksek olmasıdır. Fakat imge veri setinin boyutu çok büyük değilse, kabul edilebilir sınıflandırma performansı için birkaç evrişimsel ağ yeterli olabilir. Örneğin, Mallat ve Bruna [20,21] scattering ağı (ScatNet) adlı bir yöntem önermişlerdir. Önerilen yöntemde evrişimsel katman, doğrusal olmayan katman, pooling katmanı, karmaşık dalgacık, mod operatörü ve ortalama operatörü kullanılmıştır. Zeng ve ark. [22], tensör nesne sınıflandırması için çok çizgili diskriminant analiz ağı (MLDANet) adlı bir yöntem önermiştir. Li ve ark. [23] Kırmızı, Yeşil, Mavi, Derinlik (RGBD) imgeleri kullanarak insan hareketi tanıma için SAE-PCA yöntemini önermiştir. Son olarak, Chan ve ark. [24] temel bileşenler analiz ağı (PCANet) adlı bir yöntem önermiştir. PCANet basit işlemleri

kullanmasına rağmen, nesne görüntüleri, doku görüntüleri, el yazısı rakamları ve yüz görüntüleri ile sınıflandırmada yüksek performans sağlamıştır. Feng ve ark. [25], yabancı sahne sınıflandırması için DLANet'i (ayrıt edici lokal uyum ağını) önermişlerdir. Gan ve ark. [26] yüz tanıma için GENet'i (derin grafik gömme ağı) önermişlerdir. GENet, grafik gömme çerçevesini çok katmanlı ağ mimarisi ile birleştirmiştir. Qin ve ark. [27] kombine PCANet'i uzaysal piramit havuzu ile kullanarak balık tanıma için bir yöntem önermiştir. Lei ve ark. [28], derin tanımlayıcı yüz özelliklerine yüz tanıma için derin yüz tanımlayıcıları geliştirmişlerdir. Zhao ve ark. [29], çok düzeyli bir filtreleme ağı vasıtasıyla yerel kenarların özellik gösterimi için çok seviyeli modifiye sonlu Radon dönüşümü ağı (MMFRTN) önermiştir. Zeng ve ark. [30] PCANet'i (Temel bileşen analiz ağı) geliştirdiler. Bu yöntem, bazı veri setlerinde en yüksek sınıflandırma doğruluğunu sağlamaktadır. Derin öğrenme yoluyla özelliklerin azaltılması, akademik çalışmaların en sıcak konularından birisi haline gelmiştir.

3. ADD-TDA Tabanlı Algısal Özet Fonksiyonu

Algısal özet fonksiyonu, çoklu ortam veri işleme alanlarından birisidir ve imgelerde sıklıkla kullanılmaktadır. Algısal özet fonksiyonunun temel amacı verinin belirgin özelliklerini çıkarmaktır. Algısal özet fonksiyonu işlemi imge tanımlama, imge kimliklendirme, imge indekisleme, imge kopyalama sahteciliği tespiti vb. birçok alanda kullanılmaktadır [31]. Özellikle imge kimliklendirme işlemlerinde algısal özet fonksiyonunun sıkıştırma ve geometrik ataklara karşı dayanıklı olması istenmektedir. Kısacası, imge kimliklendirmek için kullanılacak algısal özet fonksiyonunun temel amacı, insan görme sisteminin algılamayacağı saldırılara karşı dayanıklı ve imgeyi kimliklendirebilecek bir özet fonksiyonunun oluşturulmasıdır [31-33].

Bu çalışmada, imgenin belirgin özelliklerini elde edebilmek için ADD (Ayrık Dalgacık Dönüşümü) ve TDA (Tekil Değer Ayrışımı) tabanlı yeni bir algısal özet fonksiyonu önerilmiştir. Bu yöntemde, bilineer interpolasyon uygulanarak imge yeniden boyutlandırılır ve ardından renkli imge gri seviyeye

dönüştürülür. Gri seviyeye dönüştürülen imgeye iki seviyede 2 boyutlu ayrık dalgacık dönüşümü uygulanarak Düşük-Düşük 2 (LL2) bandı elde edilmektedir. LL2 bandı elde edildikten sonra bu bant 2 x 2 boyutunda örtüşmeyen bloklara ayrılır ve her bir bloğa TDA uygulanır. TDA sonucunda tekil değerler elde edilir. Elde edilen tekil değerler normalize edilerek algısal özet değeri elde edilir. Bu yöntemde 2 seviyede ADD'nin uygulanmasının sebebi, sıkıştırma ataklarına dayanıklılığın sağlanması, TDA'nın uygulanması ise geometrik ataklara karşı dayanıklılığın sağlanmasıdır. Önerilen algısal özet fonksiyonunun blok diyagramı Şekil 1'te verilmiştir.

Önerilen algısal özet fonksiyonunun adımları aşağıdaki gibidir.

Adım 1: Bilineer interpolasyon kullanılarak imge 256 x 256 boyutunda boyutlandırılır.

Adım 2: Renkli imge gri seviyeye dönüştürülür.

$$gray = R * 0.2989 + G * 0.5870 + B * 0.1141$$

$$R = im(:, :, 1), G = im(:, :, 2), B = im(:, :, 3)$$

Adım 3: 2 boyutlu 2 seviyede ADD uygulanarak LL2 bandı elde edilir.

$$[LL \ LH \ HL \ HH] = dwt2(gri, 'filter')$$

$$[LL2 \ LH2 \ HL2 \ HH2] = dwt2(LL, 'filter')$$

Adım 4: LL2 bandı 2 x 2 örtüşmeyen bloklara ayrılır.

Adım 5: Her bir bloğa TDA uygulanarak tekil değerler elde edilir. S1 değeri algısal özet değeri olarak kaydedilir.

$$SVD(LL2_{i:i+1, j:j+1}) = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

$$S_{max} = s_1$$

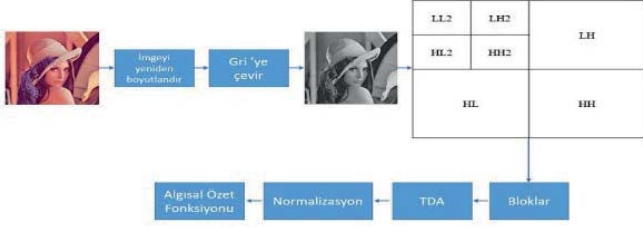
$$PH_{k,l} = S_{max}$$

$$i=\{1,2,\dots,W\}, j=\{1,2,\dots,H\}, k=\{1,2,\dots,W/2\}, l=\{1,2,\dots,H/2\}$$

Adım 6: Minimum ve maksimum değerler Pmin ve Pmax elde edilir.

Adım 7: PH matrisi normalize edilerek özet değeri elde edilir.

$$PH = round\left(\frac{PH - P_{min}}{P_{max} - P_{min}} (2^8 - 1)\right)$$



Şekil-1. Önerilen algısal özet fonksiyonunun blok diyagramı

Algoritma 1. Önerilen algısal özet fonksiyonunun sözde kodu.

```

Giriş: Im, w x h boyutlarındaki imge
Çıkış: PH, 32 x 32 boyutundaki Algısal Özet Fonksiyonu
1: İmge boyutlandır 256 x 256
2: Denklem göre gri 'ye çevir;
3: [m n]=size(LL2);
4: row=0;
5: for i=1 to m step by 2 do
6:   col=0;
7:   for j=1 to n step by 2 do
8:     [U S V]=SVD(LL2(i:i+1,j:j+1));
9:     ph(row+1,col+1)=S(1,1);
10:    col=col+1;
11:  endfor
12:  row=row+1;
13: endfor
14: PH = round( $\frac{PH-P_{min}}{P_{max}-P_{min}} (2^8 - 1)$ )

```

Bu makalede 256 x 256 x 3 boyutundaki bir imgeden 32 x 32 boyutunda algısal özet değeri elde edilmiştir ve önerilen yöntemle her bir imge için 32 x 32 x 8 =8192 bitlik özet değeri üretilmektedir. Bu özet değeri belirgin özellikleri barındırmaktadır. Bu makalede, elde edilen özet değerleri kullanılarak derin öğrenme tabanlı tanıma işlemi yapılmaktadır.

4. Derin öğrenmede ESA mimarisi

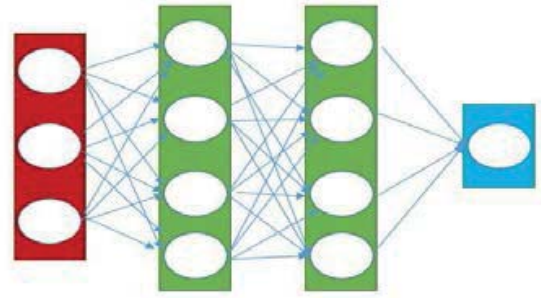
Evrişimsel sinir ağı kullanılarak imge sınıflandırma yöntemi imgelerde kullanılan en etkili derin öğrenme yaklaşımlarındandır.

3.1. Evrişimsel Sinir Ağı

ESA, tamamen bağlı bir katmanı takip eden birçok evrişimsel ve alt örnekleme katmanlarından oluşmaktadır. Bir evrişim katmanının girişinde m x m x r boyutundaki bir imge olduğunu varsayarsak; burada m, imgenin yüksekliği ve genişliği anlamına gelmektedir. Kanal sayısını ise r temsil etmektedir. Örneğin bir RGB imgesi için r = 3'tür. Evrişim katmanları için k adet n x n x q boyutundaki filtreler (veya çekirdekler) örnek verilirse; burada n imge boyutundan (m) daha küçüktür, q ise kanal sayısından

(r) küçük veya eşit olabilir ve her çekirdek için değişiklik gösterebilir. Özellik haritası üretmek için k tane filtre uygulanacaktır. Burada kullanılacak filtrelerin boyutu özellik haritası niteliği için önemlidir.

ESA yapısında bulunan öğrenilebilir ağırlıklar, ön ağırlıklar ve nöronlardan dolayı yapay sinir ağı yapısına benzemektedir. Klasik sinir ağlarından farkı, özneliklerini ham imgeden kendisinin öğrenmesidir [34]. Klasik Yapay Sinir Ağları, bir girdi alır birtakım gizli katmandan geçirdikten sonra çıkış katmanına verir. Gizli katmandaki nöronlar tamamen bağımsızdır ve birbirleriyle bağlantıları yoktur. Son katman "çıkış katmanı" olarak adlandırılır ve sınıflandırma işlemindeki doğruluk oranı bu katman çıktısı hesaplanarak elde edilir [35]. Klasik sinir ağı yapısı şekil 2'de verilmiştir.



Giriş Katmanı 1. Gizli Katman 2. Gizli Katman Çıkış Katmanı

Şekil-2: Klasik sinir ağı yapısı

Klasik sinir ağları ham imgeleri sınıflandıramaz. Örneğin, 32x32x3'lük bir renkli imgede, ilk gizli katmanda 32 * 32 * 3 = 3072 ağırlığı olacaktır. Bu miktar, klasik sinir ağlarının üstesinden gelebileceği bir problemdir. Fakat 227x227x3 boyutundaki bir imgede, gizli katmanda 227 * 227 * 3 = 154578 ağırlığı olacak ve aşırı uyumluluk problemini ortaya çıkartacaktır.

3.1.1. ESA yapısında kullanılan katmanlar

Evrişimsel katmanlar, Giriş – Evrişim katmanı (Convolution) - Doğrultulmuş doğrusal birim (Rectified Linear Unit (ReLU)) – Pooling - Tam bağlı katmanlardan (Fully Connected Layer) oluşmaktadır.

ReLU: Aşağıda ReLU aktivasyon fonksiyonunun formülü verilmiştir.

$$y_{ijk} = \max\{0, x_{ijk}\} \quad (1)$$

Relu fonksiyonu tüm negatif durumları sıfıra dönüştürür; bu da ağıın doğrusal olmayan özelliklerini arttırır.

Pooling: Relu katmanlarından sonra yapılır. ESA mimarisinde, pooling katmanı, evrişim katmanları arasında periyodik olarak eklenebilir. Pooling işlemi evrişim katmanından sonra öznitelik boyutunu azaltmaktadır. Pooling ile evrişim katmanlarından sonra boyutun kademeli olarak azaltılmasının amacı, ağdaki parametre ve hesaplama miktarını azaltmaktır, böylece aşırı uyumluluk üzerine bir kontrol sağlanır. Pooling yapmanın en yaygın yolu, denklem 2'de gösterilen maksimum pooling'dir.

$$y_{ijk} = \max\{y_{i'j'k'}: i \leq i' < i+p, j \leq j' < j+q\} \quad (2)$$

Özellik haritası: Her katmandan sonra özellik haritası için aşağıdaki formül uygulanır. Kullandığımız ESA mimarisine göre, giriş katındaki 227x227x3 boyutlu imgeler tek katmandan sonra evrişilerek 55x55x96 özellik haritası elde edilir [36].

$$\text{Boyut}_{\text{oh}} = \mathcal{X} = \left(\frac{\text{size}_{-x_i} - F_i + 2\text{Pad}}{\text{Stride}} \right) + 1 \quad (3)$$

Kullandığımız ESA mimarisine göre katman 1'de kullanılan parametreler aşağıdaki gibidir.

1. Katman: 96 filtreli evrişim, 11x11 boyutundaki evrişim matrisi, her evrişim matrisinden sonra 4 adım kayma (stride) 4 ve çıkan matrisin etrafını 0 ile doldurma (padding) 0.

Denklem 3'teki formülü uyguladığımızda; $(227-11)/4 + 1 = 55$ elde ederiz. Bu sonuç ilk katmandan sonra çıkan özellik kümesinin boyutunu temsil eder. Dolayısıyla, birinci katmanda 96 filtre uygulandığı için 55x55x96 boyutunda bir özellik kümesi elde edilir. Uygulanan mimari adımları Çizelge-2'deki gibidir.

Çizelge-1: Uygulanan ESA mimarisinin adımları

Adım Sırası	Katman adı	Yapılan işlem
1	Evrişim Katmanı	96 11x11x3 evrişim, stride [4 4] ve padding [0 0].
2	Maks Pooling	3x3 maks pooling, stride [2 2] ve padding [0 0].
3	Evrişim Katmanı	256 5x5x48 evrişim, stride [1 1] ve padding [2 2].
4	Maks Pooling	3x3 maks pooling, stride [2 2] ve padding [0 0].
5	Evrişim Katmanı	384 3x3x256 evrişim, stride [1 1] ve padding [1 1].
6	Evrişim Katmanı	384 3x3x192 evrişim, stride [1 1] ve padding [1 1].
7	Evrişim Katmanı	256 3x3x192 evrişim, stride [1 1] ve padding [1 1].
8	Maks Pooling	3x3 maks pooling, stride [2 2] ve padding [0 0].
9	Tam Bağlı Katman	Öznitelikler

5. Önerilen ADD-TDA-ESA yöntemi

Bu makalede, ADD-TDA tabanlı algısal özet fonksiyonu kullanılarak imgeler 32x32 boyutuna indirgenmiştir. Orijinal imgeye, 2. seviyede 2 boyutlu ayrık dalgaçık dönüşümü uygulanarak 2 x 2 boyutunda örtüşmeyen bloklara ayrılmıştır. Daha sonra ham imgelerden belirgin özelliklerin çıkarılması için her bir bloğa TDA uygulanmıştır. TDA sonucunda elde edilen tekil değerler normalize edilerek algısal özet değeri elde edilmiştir. Önerilen ADD-TDA algısal özet fonksiyonu, Caltech-101 veri tabanından 7 sınıfa (uçak, araba, araba arka planı, yüzler, gitarlar, evler, motosikletler) toplam 3150 imgeye uygulanmıştır. Bu 3150 imge algısal özet fonksiyonunun sonucu olarak 32x32 boyutundaki imgelere dönüştürülüp ESA mimarisine girdi olarak verilip ve SVM ile sınıflandırılmıştır.

Önerilen ADD-TDA-ESA yöntemi, derin öğrenmenin önde gelen mimarilerinden Alexnet mimarisi ile kullanılmıştır. Alex Krizhevsky, Ilya Sutskever ve Geoffrey Hinton tarafından geliştirilen mimari, 2012 Resim Sınıfı Yarışması'nı (ILSVRC) kazanmıştır [34]. Bu makalede öznitelik çıkarımı için Alexnet'in ESA mimarisinde bulunan ve birbirlerini takip eden beş evrişim katmanı kullanılmıştır. Daha sonra bu katmanlardan elde edilen ve her bir imgeye ait 4096 öznitelik, destek vektör makinesine verilerek sınıflandırılmıştır. Önerilen yöntem Şekil-3'te gösterildiği gibidir.

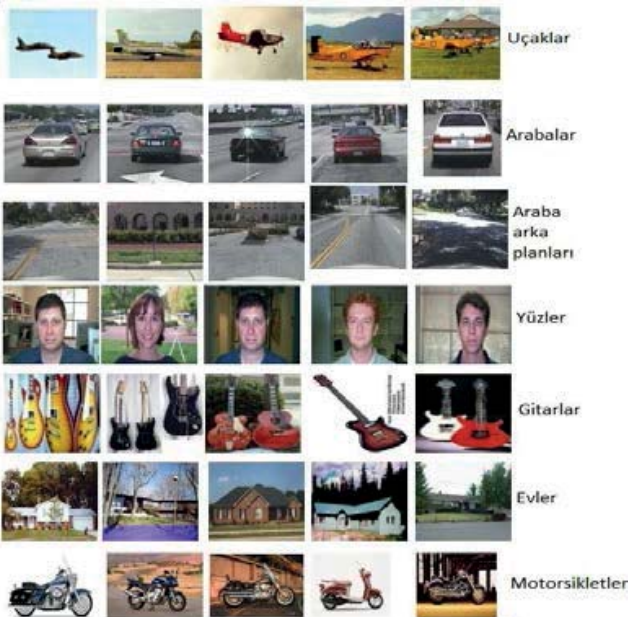


Şekil-3: Önerilen ESA mimarisi

Veri Seti; Caltech-101 veri setinde 101 farklı kategoride çeşitli imgeleri bulunur; Fei-Fei ve arkadaşları tarafından Google resim arama ile toplanan bir arka plan kategorisi de vardır [37]. Bu makalede, ADD-TDA-ESA yöntemi; Caltech-101 imge veri tabanında bulunan uçaklar, arabalar, araba arka planları, yüzler, gitarlar, evler ve motosikletlerden oluşan 7 farklı sınıfa uygulanmaktadır. ADD-TDA-ESA yönteminde kullandığımız 7 sınıfın örnek gösterimi Şekil-4 ve 5'te gösterilmiştir.



Şekil-5: Algısal Özet Fonksiyonu tarafından elde edilen imgeler.



Şekil-4: Caltech 101 veri tabanından örnek imgeler

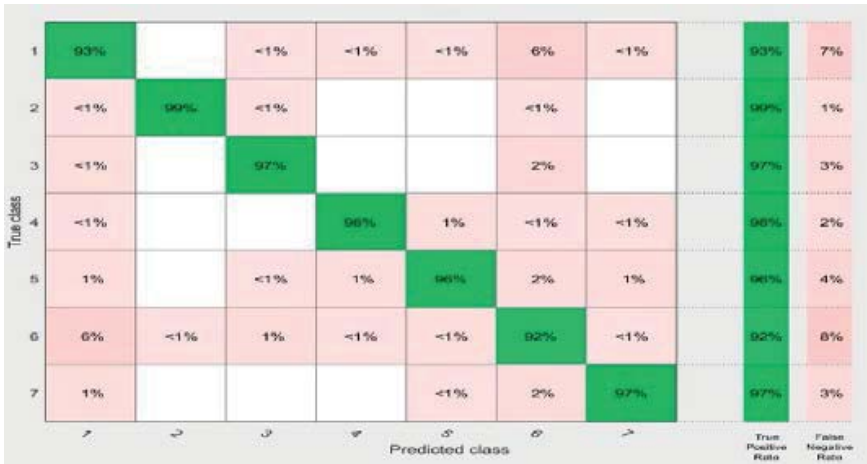
6. Deneysel Sonuçlar

Önerilen ADD-TDA-ESA yöntemi, Intel Core i7 - 4510U işlemci, 8 GB RAM ve Windows 10 işletim sistemli bir dizüstü bilgisayarda gerçekleştirildi. Uygulamanın kodları MATLAB R2016a'da MatConvnet Kütüphanesi kullanılarak yazılmıştır. Caltech-101 veri tabanına ait imgelerin boyutları 400x600x3 ile 280x186x3 arasında değişmektedir. Her sınıfa ait 450 imge kullanılmıştır. Bu imgelerin %70'i eğitim, %30'u test olarak kullanılmıştır. Algısal özet fonksiyonu kullanılmadan elde edilen başarı oranı %96,1, çalışma süresi ise 241,21 saniyedir.

Her imgenin 32 x 32 x 8 = 8192 bit 'ten oluşan belirgin özellikleri ADD-TDA tabanlı algısal özet fonksiyonu kullanılarak elde edilmiştir. Bu özet değer 32x32 boyutlu imgeye dönüştürülerek, %70'i eğitim, %30'u test olarak kullanılmıştır. Bu imgelerin sınıflandırma başarı oranı %95,8, çalışma süresi 83,08 saniyedir. Her iki yöntemin çalışma süreleri Çizelge-3'te verilmiştir. Klasik yöntem ile önerilen yöntemin karışıklık matrisleri Şekil-6 ve 7'de verilmiştir.

Çizelge-3: Önerilen yöntemin karşılaştırılması

Yöntemler	Sınıflandırma doğruluğu (%)	Yürütme Süreleri
Klasik ESA	96,1	241,21 saniye
ADD-TDA-ESA	95,8	83,08 saniye



Şekil-6: Klasik ESA karışıklık matrisi



Şekil-7: Önerilen yöntemin karışıklık matrisi

7. Sonuç ve Gelecek Çalışmalar

Bu makalede, imge sınıflandırması için algısal özet fonksiyonu tabanlı yeni bir derin öğrenme mimarisi sunulmuştur. Önerilen ADD-TDA-ESA yöntemi, imge sınıflandırma süresini azaltırken sınıflandırma performansını kabul edilebilir bir seviyede tutmuştur. Bu yöntem, algısal özetleme, evrimsel özellik çıkarma ve sınıflandırma olmak üzere 3 adımdan oluşmuştur. ADD-TDA tabanlı algısal özet, önerilen ADD-TDA-ESA yönteminde kullanılmıştır. Bu aşamaya ön işleme safhası da denilebilir. İkinci aşamada, ESA'nın evrimsel yöntemi kullanılarak bir özellik kümesi elde edilmiştir. Elde edilen özellikler destek vektör makinesi kullanılarak sınıflandırılmıştır. Sınıflandırmadaki doğruluk oranı ve çalışma süresi parametreleri, önerilen yöntemin performansını ölçmek için kullanılmıştır. Ayrıca, önerilen algısal özet fonksiyonunun güçlü özellikleri sayesinde, 32x32 boyutundaki küçük imgeler, Caltech-101 veri tabanındaki boyutları 400x600x3 ile 280x186x3 arasında değişen imgelere kıyasla tatmin edici sınıflandırma sonucu verirken çalışma süresini yaklaşık olarak üçte bir oranında azaltmıştır. Çalışmada kullanılan ESA mimarisi ve algısal özet fonksiyonu değiştirilip daha kapsamlı çalışmalar yapılabilir. Önerilen yöntem, diğer gelişmiş derin öğrenme ağlarının çalışma sürelerinin incelenmesi için değerli temel bir başlangıç noktası olabilir.

Kaynakça

- [1] Lowe, D. G. (2004). Distinctive image features from scale-invariant key points. *International journal of computer vision*, 60(2), 91-110.
- [2] Jain, A. K., Ratha, N. K., & Lakshmanan, S. (1997). Object detection using Gabor filters. *Pattern recognition*, 30(2), 295-309.
- [3] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7), 971-987.
- [4] J. Virmani, V. Kumar, N. Kalra, N. Khandelwal, SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors, *J. Digit. Imaging* 26 (3) (2012) 530–543.
- [5] N.K. Jitendra Virmani, Vinod Kumar Naveen Kalra, Prediction of liver cirrhosis based on multiresolution texture descriptors from B-mode ultrasound, *Int. J. Converg. Comput.* 1 (2013) 1–19.
- [6] U.R. Acharya, H. Fujita, S. Bhat, U. Raghavendra, A. Gudigar, F. Molinari, A. Vijayanathan, K. Hoong Ng, Decision support system for fatty liver disease using GIST descriptors extracted from ultrasound images, *Inf. Fusion* 29 (2016) 32–39.
- [7] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
- [8] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.
- [9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [10] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, 2(1), 1-127.
- [11] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- [12] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27-48.
- [13] Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267-285). Springer Berlin Heidelberg.
- [14] A. Ng, Sparse autoencoder, CS294A Lecture Notes, vol. 72, 2011
- [15] Salakhutdinov, R., & Hinton, G. (2009, April). Deep Boltzmann machines. In *Artificial Intelligence and Statistics* (pp. 448-455).
- [16] Sutskever, I., Hinton, G. E., & Taylor, G. W. (2009). The recurrent temporal restricted Boltzmann machine. In *Advances in Neural Information Processing Systems* (pp. 1601-1608).
- [17] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- [18] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [19] LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-104). IEEE.
- [20] Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10), 1331-1398.
- [21] Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1872-1886.
- [22] Zeng, R., Wu, J., Senhadji, L., & Shu, H. (2015, April). Tensor object classification via multilinear discriminant analysis network. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 1971-1975). IEEE.
- [23] Li, S. Z., Yu, B., Wu, W., Su, S. Z., & Ji, R. R. (2015). Feature learning based on SAE-PCA network for human gesture recognition in RGBD images. *Neurocomputing*, 151, 565-573.
- [24] Chan, T. H., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification?. *IEEE Transactions on Image Processing*, 24(12), 5017-5032.
- [25] Feng, Z., Jin, L., Tao, D., & Huang, S. (2015). DLANet: A manifold-learning-based discriminative feature learning network for scene classification. *Neurocomputing*, 157, 11-21.
- [26] Gan, Y., Yang, T., & He, C. (2014, October). A deep graph embedding network

- model for face recognition. In Signal Processing (ICSP), 2014 12th International Conference on (pp. 1268-1271). IEEE.
- [27] Qin, H., Li, X., Liang, J., Peng, Y., & Zhang, C. (2016). DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, 187, 49-58.
- [28] Lei, Z., Yi, D., & Li, S. Z. (2016). Learning stacked image descriptor for face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(9), 1685-1696.
- [29] Zhao, Y., Wang, R., Wang, W., & Gao, W. (2016). Multilevel modified finite radon transform network for image upsampling. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(12), 2189-2199.
- [30] Zeng, R., Wu, J., Shao, Z., Chen, Y., Chen, B., Senhadji, L., & Shu, H. (2016). Color image classification via quaternion principal component analysis network. *Neurocomputing*, 216, 416-428.
- [31] Tang, Z., Zhang, X., Dai, X., Yang, J., & Wu, T. (2013). Robust image hash function using local color features. *AEU-International Journal of Electronics and Communications*, 67(8), 717-722.
- [32] Qin, C., Chang, C. C., & Tsou, P. L. (2013). Robust image hashing using non-uniform sampling in discrete Fourier domain. *Digital Signal Processing*, 23(2), 578-585.
- [33] Tang, Z., Zhang, X., & Zhang, S. (2014). Robust perceptual image hashing based on ring partition and NMF. *IEEE Transactions on Knowledge and Data Engineering*, 26(3), 711-724.
- [34] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [35] Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1), 3-31.
- [36] Gao, X., Li, W., Loomes, M., & Wang, L. (2017). A fused deep learning architecture for viewpoint classification of echocardiography. *Information Fusion*, 36, 103-113.
- [37] Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer vision and Image understanding*, 106(1), 59-70.