Kapsül Ağları ile İşaret Dili Tanıma
Recognition of Sign Language using Capsule Networks

Fuat BEŞER_1, Merve Ayyüce KIZRAK_2, Bülent BOLAT_3 ve Tülay YILDIRIM_4
Elektronik ve Haberleşme Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
e-mail fuatbeser@gmail.com_1, e-mail ayyucekizrak@gmail.com_2, e-mail bbolat@ytu.edu.tr_3, e-mail tulay@ytu.edu.tr_4

***Özetçe*— İşitme ve konuşma engelliler, dudak okuma ya da el ve yüz hareketlerinden oluşan ifadeler yardımıyla iletişimlerini sürdürmektedirler. Engelli bireylerin topluma katılımlarının sağlanması ve yaşam kalitelerinin artırılması diğer insanlarla sağlıklı ve etkili bir şekilde iletişim kurmaları ile mümkün olmaktadır. Bu çalışmada; işaret diline ait rakamların, derin yapay sinir ağı (deep artificial neural network) modeli olan Kapsül Ağları ile %94,2 başarı ile tanınması sağlanmıştır.**

Anahtar Kelimeler — derin öğrenme; derin sinir ağları; kapsül ağları; evrişimli sinir ağları; işaret dili; işaret dili tanıma.

***Abstract*— Hearing and speech impaired persons continue to communicate with the help of lip reading or hand and face movements also known as a sign language. Ensuring the disabled persons participation in life and increasing their quality of life are achievable through healthy and effective communication with other people. In this work; digits of the sign language were recognized with %94.2 validation accuracy by Capsule Networks.**

Keywords — deep learning; deep neural networks; capsule networks; convolutional neural networks; sign language; sign language recognition.

I. GİRİŞ

Derin öğrenme ile ilgili çalışmalar, Toronto Üniversitesi'nden Geoffrey Hinton, Ilya Sutskever ve Alex Krizhevsky'nin, 2012 yılında 1000 farklı sınıfa ait nesnelerin tanınmaya çalışıldığı "ImageNet Large Scale Visual Recognition Challenge" isimli yarışmayı AlexNet [2] isimli derin evrişimli sinir ağı (deep convolutional neural network) mimarisi ile -en iyi 5 sonuç dikkate alındığında- en yakın rakibine kıyasla %10,9 daha az hata yaparak kazanması ile dünya çapında hız kazanmıştır. Hinton ve ekibi yarışmayı kazanabilmek için 60 milyon parametreye, 650 bin nörona ve beşi evrişim (convolution) ve üçü de tam bağlantılı olmak üzere toplam sekiz katmana sahip AlexNet'i 1,2 milyon yüksek çözünürlüklü resim ile eğitmiştir.

Her ne kadar nesne tanımada evrişimli sinir ağları ile umut vadeden sonuçlar elde ediliyor olsa da evrişimli sinir ağları bazı problemleri beraberinde getirmektedir. Eğitilmiş evrişimli sinir ağı, üzerinde eğitilmiş olduğu nesneye farklı bir perspektiften bakıldığında aynı başarı ile tanıma işlemini gerçekleştirememektedir. Evrişimli sinir ağının bir nesneyi oluşturan parçaların arasındaki hiyerarşiyi (örneğin bir yüzün göz, ağız, burun vb. organlardan oluşması) anlayamadığı bilinen bir gerçektir. Yapay genel zekaya (artificial general intelligence) giden yolda nesnelerin konum, yönelim ve açısal durumdan bağımsız olarak tanınabilmesi için farklı parametrelerle temsil edilmesi gerekmektedir. Bakış açısından bağımsız nesne tanıma [4], özelleştirilmiş bir evrişimli sinir ağı mimarisi olarak kabul edilebilen kapsül ağı ve kapsüllerin aralarındaki dinamik bağlantılar sayesinde mümkün olmaktadır [1]. Kapsül ağlarıyla ilgili bugüne kadar yapılan çalışmalar Tablo 1'de yer almaktadır.

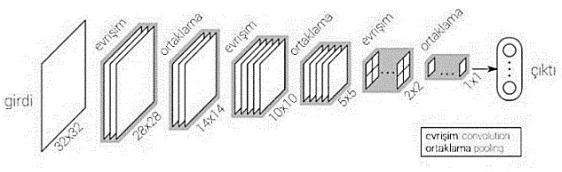
TABLO I. ÖNCEKİ ÇALIŞMALARDA ELDE EDİLEN BAŞARIMLAR

Çalışmayı Gerçekleştirenler	Çalışmanın İçeriği		
	Veri Seti	*Giriş Resim Boyutu*	*Test Başarısı*
Sabour vd. [1]	MNIST [5]	28 x 28 piksel	%99,75
Xi vd. [2]	CIFAR 10 [6]	32 x 32 piksel	%71,55
Qiao vd. [11]	MNIST [5] ve FMRI	28 x 28 piksel	%75

Bu çalışmada kapsül ağı ve dinamik yönlendirme algoritması kullanılarak işaret diline ait görüntülerin sınıflandırılması üzerine çalışılmıştır; 2. Bölümde kapsül ağı ve dinamik yönlendirme algoritması, 3. bölümde kullanılan yöntem, yazılım ve donanımlar, 4. Bölümde üzerinde çalışma yapılan veri seti ve özellikleri belirtildikten sonra 5. Bölümde uygulamanın avantaj ve dezavantajları, seçilmesi gereken parametreler ve elde edilen sonuçlar karşılaştırmalı olarak değerlendirilmektedir.

II. KAPSÜL AĞLARINA GENEL BİR BAKIŞ

Temel konsepti Hubel ve Wiesel [10] tarafından yapılan deneylerle ortaya konulan, Fukushima [8] tarafından modellenen ve Lecun vd. [5] tarafından ilk başarılı uygulaması geliştirilen evrişimli sinir ağları, başarılı sonuçlar vermesi nedeniyle bilgisayarlı görü uygulamalarında sıklıkla tercih edilmektedir.



Şekil. 1. Evrişimli Sinir Ağı Modeli

978-1-5386-1501-0/18/$31.00 ©2018 IEEE

Evrişimli sinir ağı modelindeki evrişim katmanlarının çıkışlarında, görüntüye dair öznitelik haritaları (feature map) elde edilmektedir. İlk katmanlarda, kenar gibi daha basit bilgileri içeren öznitelikler tespit edilirken; daha üst seviyeli katmanlarda, ilk katmanlarda elde edilen öznitelikler kullanılarak görüntünün geneline ilişkin daha karmaşık öznitelikler çıkarılmaktadır. Evrişimli sinir ağları, ileri yönlü bir model olup insan görme sisteminden esinlenerek tasarlanmıştır. Evrişim işlemi, bir sinir hücresinin kendi uyarı alanında uyaranlara verdiği tepkinin/yanıtın modellenmesi olarak düşünülebilir [8]. Genellikle evrişim uygulanan katmanların çıkışlarında oluşan tensörlerin genişlik ve yüksekliği azalırken, kanal sayısı kullanılan filtre sayısına bağlı olarak değişmektedir. Artan kanal sayısı hesaplama yükünü artırdığından evrişim katmanlarının yanı sıra boyut azaltmak ve eğitim ile test başarı oranları arasındaki farkı minimum yapmak için Şekil 2'de gösterildiği gibi ortalama ortaklama (average pooling) ve maksimum ortaklama (max-pooling) katmanları da kullanılmaktadır.

Şekil. 2. Maksimum ve Ortalama Ortaklama

Ortaklama işlemi vasıtasıyla gerçekleştirilen boyut azaltma işlemi aynı zamanda bilgi kaybına neden olmaktadır. Ayrıca tanınmaya çalışılan nesnelerin örtüşüyor (overlapping) olması da evrişimli sinir ağı modellerinin sınıflandırma ve bölütleme yaparken zorlanmasına neden olmaktadır [1].

Evrişimli sinir ağı modellerinin yetersiz kaldığı problemlere çözüm olarak kapsül ağları ve dinamik yönlendirme algoritması [1, 3] önerilmiştir. Konum, yönelim, duruş ve açısal değerin değişmesi durumunda dahi bir grup nörondan oluşan kapsüller aracılığıyla nesnenin başarıyla tanınabilmesi için nesneyi temsil eden kalınlık, ölçek, kaydırma vb. özelliklerin anlaşmalı yönlendirme (routing-by-agreement) ile öğrenilmesi önerilmiştir [1].

A. Dinamik Yönlendirme

Kapsül modelinde standart bir sinir ağı modelinden farklı olarak kapsül katmanlarının çıkışları skaler değil vektörel olarak ifade edilmektedir. Derin sinir ağları modellerinde tek bir nöron için çoğunlukla doğrusal olmayan ReLU (Rectified Linear Unit) aktivasyon fonksiyonu kullanılırken, kapsül ağlarında girdisinin ve çıktısının vektör olduğu denklem (1)'de gösterilen ezme (squash) fonksiyonu kullanılmaktadır.

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2}\frac{s_j}{\|s_j\|} \tag{1}$$

Ezme işlemi klasik modellerde kullanılan maksimum ortaklama işleminden çok daha güçlü özniteliklerin oluşmasını sağlar. Ancak dinamik yönlendirme işlemi fazladan hesaplama maliyeti gerektirir. Bir sonraki katman için ağırlıkların toplamı denklem (2)'de ve afin dönüşümü denklem (3)'te gösterildiği şekilde hesaplanmaktadır.

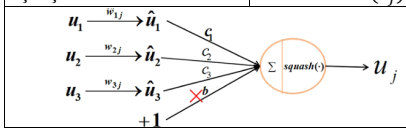
$$s_j = \sum_i c_{ij}\hat{u}_{j|i} \tag{2}$$

$$\hat{u}_{j|i} = W_{ij}.u_i \tag{3}$$

Kapsül ağında hesaplanan vektörler, nesneleri temsil eden konum, yönelim, kalınlık ve yönlendirme gibi öznitelikleri belirtmektedir. Görüntüde nesnenin bulunmadığı alanlarda vektör değerleri küçük olurken, nesne tespit edilen alanlarda vektörün boyutları özniteliğe bağlı olarak değişmektedir. Denklem (4) ile esnek/türevlenebilir eşikleyici (softmax) hesabı ifade edilmektedir.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{4}$$

TABLO II. KAPSÜL AĞI İLE KLASİK SİNİR AĞI MODELİ KARŞILAŞTIRMASI [9]

		Kapsül Ağı	**Klasik Sinir Ağı**
Düşük Seviyeli Sinir Hücresi/Kapsül		***vektörel***	***skaler***
İşlem	Afin Dönüşümü	$\hat{u}_{j\|i} = W_{ij}.u_i$	-
	Ağırlıklandırma/ Toplama	$s_j = \sum_i c_{ij}\hat{u}_{j\|i}$	$a_j = \sum_i W_i x_i + b$
	Doğrusal Olmayan Fonksiyon	$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2}\frac{s_j}{\|s_j\|}$	$h_{W,b}(x) = f(a_j)$
Çıkış		*Vektör* (v_j)	*Skaler (h)*
			

Kapsül ağlarında evrişim katmanının çıkışı öncül kapsüle giriş olarak uygulanmaktadır. Çıkışlardan vektörler elde edebilmek için yeniden boyutlandırma yapılmaktadır. Bu vektörlerin değerlerini sınırlandırabilmek için ezme işlemi uygulanır. Algoritma aşağıda belirtildiği gibidir:

Algoritma 1: Yönlendirme [1].

1. **algoritma** Yönlendirme $(\hat{u}_{j|i}, r, l)$
2. tüm kapsüller için i. katman l ve kapsül j. katman $(l+1): b_{ij} \leftarrow 0$
3. **for** r iterasyon **do**
4. tüm kapsüller için i. katman $l: \boldsymbol{ci} \leftarrow softmax(\boldsymbol{b}_i) \Rightarrow$ softmax denklemi (4)
5. tüm kapsüller için j. katman $(l+1): \boldsymbol{s}_j \leftarrow \sum_i c_{ij}\hat{\boldsymbol{u}}_{j|i}$
6. tüm kapsüller için j. Katman $(l+1): v_j \leftarrow ezme(\boldsymbol{s}_j) \Rightarrow$ ezme denklemi (2)
7. tüm kapsüller için i. katman l ve kapsül j. katman $(l+1): b_{ij} \leftarrow b_{ij} + \hat{\boldsymbol{u}}_{j|i}.\boldsymbol{v}_j$

return v_j

978-1-5386-1501-0/18/$31.00 ©2018 IEEE

Aktivasyon haritası olarak adlandırılan kapsül çıkışlarında oluşan vektörlerin uzunlukları ve yönleri, nesneyi temsil eden özniteliklere ait bilgileri içermektedir [1]:

- Vektör uzunluğu: Nesnenin o noktadaki tahmini bulunma olasılığını temsil eder.
- Vektör yönü: Nesnenin poz, deformasyon, hız, yansıtabilirlik (albedo), renk derecesi (hue) ve doku (texture) parametrelerini verir.

B. Kapsül Ağlarının Avantaj ve Dezavantajları

Kapsül ağında, evrişimli sinir ağı modelinden farklı olarak kapsül yapısı, dinamik yönlendirme ve ezme fonksiyonu kullanılarak nesne tanıma, sınıflandırma ve bölütleme için daha dayanıklı bir model elde edilmektedir [1, 3, 9]. Ancak bu yeni yöntemin de avantajlarının yanında dezavantajları vardır. Tablo III'te bunlar özetlenmektedir:

TABLO III. KAPSÜL AĞLARININ AVANTAJ VE DEZAVANTAJLARI

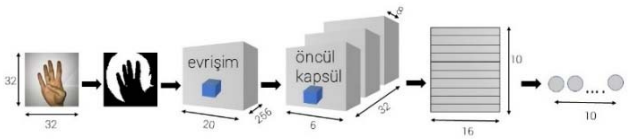
Avantajlar	Dezavantajlar
MNIST veri setinde en yüksek başarıma sahiptir. CIFAR10 veri setinde de umut vadetmektedir.	CIFAR10'da temel modellerden başarılı değildir ancak geliştirilebilir bir başlangıç sergilemektedir.
Daha küçük veri setlerinde başarılıdır.	Henüz çok büyük veri setlerinde test edilmemiştir.
Nesnenin poz, deformasyon, hız, yansıtabilirlik (albedo), renk derecesi (hue) ve doku (texture) bilgilerini korur (equivariance).	Anlaşmalı yönlendirme algoritmasından dolayı eğitim için daha fazla süre gerekmektedir.
Görüntü bölütleme ve nesne tanıma konularında umut vadetmektedir.	
Anlaşmalı yönlendirme algoritması çakışan görüntülerde nesneleri kolaylıkla ayırt etmeyi sağlamaktadır.	Kapsül ağı benzer özelliklere sahip iki nesneyi ayırt etmekte zorlanabilmektedir.
Aktivasyon vektörleri sayesinde görüntüyü yorumlamak kolaylaşmaktadır.	

III. YÖNTEM

Bu çalışmada, işaret diline ait rakamları belirtilen görüntülerde sınıflandırma işlemi kapsül ağı kullanılarak gerçekleştirilmiştir. Oluşturduğumuz modelin ilk katmanında 7×7 evrişim işlemi uygulanmaktadır. Bu işlem sonucunda 26×26×256 boyutlu tensör, öncül kapsül katmanına iletilmektedir. Kapsül katmanında yapılan işlemler ve elde edilen çıkış boyutları sırasıyla aşağıdaki gibidir;

- Evrişim: 9 × 9 × 256
- Yeniden boyutlandırma: 2592 × 8
- Ezme: 2592 × 8
- Kapsül katmanı çıkışı (DigitsCaps): 10 × 16

Öncül kapsül katmanı [32 × 6 × 6] boyutludur ve her bir çıkış vektörü 8 elemanlıdır. Her bir kapsülde [6 × 6] elemanlı bir ağırlık paylaşımı bulunmaktadır. Bu çalışma için kullanılan kapsül ağı modeli şekil 4'te olduğu gibi, girişte bir evrişim, öncül kapsül katmanı ve kapsül çıkışı olacak şekildedir.



Şekil. 4. İşaret Dili Tanıma için Kapsül Modeli

Kapsül ağı modelinin çıkış katmanında (DigitsCaps) her bir sınıf için 16 uzunluklu kapsül bir önceki katmana bağlı olarak hesaplanmaktadır. Son aşamalarda 3 tane tam bağlantılı (fully connected: FC) katmana sırasıyla aktivasyon fonksiyonu olarak ReLU, ReLU ve sigmoid uygulanmaktadır. Görüntünün tekrar oluşturulabilmesi (reconstruct) için sigmoid fonksiyonu kullanılmaktadır. Bu aşamada görüntünün tekrar oluşturma kaybı olan α değeri 0,005 olarak seçilmiştir. Esnek/türevlenebilir eşikleyici ile 10 sınıf için başarım sonucu elde edilmektedir.

Çıkış katmanında elde edilen 160 boyutlu vektör tam bağlantı katmanları ile klasik sinir ağı yapısıyla tamamlanmaktadır. Tüm model için hesaplanan ve öğrenilen toplam parametre sayısı 10,296,576 olmaktadır. Bu çalışmada uygulanan derin öğrenme modelleri ile;

- NVIDIA GTX 1080
- TESLA K80

olmak üzere iki farklı donanım üzerinde deneyler gerçekleştirilmiştir. TESLA K80 donanımı Google Colaboratory aracılığıyla ücretsiz olarak kullanılmıştır. Modeller Keras kütüphanesi kullanılarak Python programlama dilinde oluşturulmuştur.

IV. VERİ SETİ: İŞARET DİLİ İLE RAKAMLAR

Bu çalışma için rakamları ifade eden işaret dili veri seti seçilmiştir [7]. Bu veri seti, işaret dilini sınıflandırmak ve sese çevirmek amacıyla oluşturulmuştur. Bu veri seti ile gerçekleştirilecek çalışmalar sayesinde işitme ve konuşma engelli bireylerin hayat kalitesini artırmak ve topluma katılımlarını kolaylaştırmak hedeflenmiştir. 218 farklı katılımcıdan, her bir ifade için 10 örnek görüntü alınarak oluşturulmuş olan veri setindeki görüntüler 100×100 piksel boyutundadır. Üç kanallı (RGB) görüntülerden oluşan veri seti, 0-9 rakamlarını kapsayan 10 adet sınıfı içermektedir [7]. Veri setinde yer alan resimlere ön işlem olarak yeni boyut 32 x 32 ve tek kanallı olacak şekilde yeniden boyutlandırma işlemi uygulanmıştır. Veri seti %70 eğitim ve %30 test olacak şekilde ayrılmıştır.

Şekil. 5. İşaret Dili ile Rakamlar Veri Seti [7].

978-1-5386-1501-0/18/$31.00 ©2018 IEEE

Veri setinde yer alan simgelerin doğru anlaşılması önemli bir konudur. Görüntü içeriğinin konum, yönelim ve açısı simgenin anlamıyla doğrudan ilişkili olduğu için evrişimli sinir ağları ile kapsül ağları, başarım analizi yapılmak için tercih edilmektedir.

V. SONUÇLAR

Evrişimli sinir ağı modellerinde yaşanan bazı problemleri gidermesi beklenen kapsül ağları ile MNIST (Modified National Institute of Standards and Technology) veri seti üzerinde %99,75 başarı elde edilmiştir [1]. Özellikle görüntüde konum, yönelim ve açı bilgisinin diğer derin öğrenme modellerine göre daha verimli bir şekilde kullanması dolayısıyla, bu çalışmada dinamik yönlendirme algoritmasına sahip kapsül ağı modeli kullanılmıştır [1, 3, 8].

İşaret dili ile rakamlar veri setinin çeşitli veri artırma, farklı küme boyutları (batch size), evrişim katmanı eklenmesi ve filtre boyutunun etkisi gibi deneyler ile modelin performansı değerlendirilmiştir. Rastgele parametre başlangıcının etkisini sınırlandırmak ve çıkışta gerçekçi sonuçlar elde edebilmek için 50 epoch 10 kez çalıştırılmış ve başarımların ortalaması alınmıştır.

TABLO IV. KAPSÜL AĞI MODELİNE UYGULANAN YÖNTEMLERİN BAŞARIMA ETKİSİ.

Modele Uygulanan Yöntem		**Ortalama Başarım (%)**
Veri Artırma	%10 Kaydırma	93,72
	15 Derece Döndürme	91,78
	5 Derece Döndürme	91,57
	Dikey Döndürme (Yansıtma)	88,30
Küme Boyutu (Batch Size)	32	89,48
	64	87,06
	128	80,33
Evrişim Katmanı	128	90,52
	64 + 128	90,91
Filtre Boyutu	9	91,85
	7	92,10
	5	91,32
	3	89,41

Bu çalışma ile;

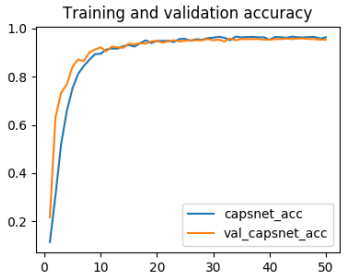
- Kapsül modelinin başarımında %10 kaydırma veri artırma tekniği ile yaklaşık %4'lük bir artış (küme boyutu 32 ve filtre boyutu 3 olduğu duruma kıyasla) sağlanmış,
- Veri setinde çok fazla görüntü olmamasından dolayı küme boyutunu artırmanın başarımı olumsuz etkilediği gözlemlenmiş,
- Kapsül katmanından önce evrişim katmanı eklenmesinin, yani modelin derinleştirilmesinin, başarımı az da olsa artırdığı anlaşılmış,

Elde edilen sonuçlar ışığında; en başarılı sonuçlardan yeni bir kombinasyon oluşturulmuştur.

TABLO V. KAPSÜL AĞI MODELİNE UYGULANAN YÖNTEMLERİN KOMBİNİZASYONU İLE BAŞARIMA ETKİSİ.

Modele Uygulanan Yöntem		**Ortalama Başarım (%)**
Veri Artırma	%10 Kaydırma	94, 2
Küme Boyutu	32	
Filtre Boyutu	7	

Bu yeni kombinasyonda test başarımının %95,4'e kadar çıktığı zamanlar olduğu tespit edilmiştir. Oluşturulan model 10 kez çalıştırılıp sonuçların ortalaması alınarak elde edilen başarım Tablo V'te ve elde edilen başarım oranı değişimi Şekil 6'da gösterilmektedir.



Şekil. 6. Eğitim ve Geçerleme (Validation) Başarım Değişimi

Sonuç olarak, literatürde henüz çok yeni olan kapsül ağı modelinin daha önce hiç denenmemiş bir veri seti üzerindeki etkisi incelenmiş ve farklı durumlardaki davranışları yorumlanmıştır. En iyi koşulların kombinasyonu ile elde edilen ortalama başarım %94,2 olarak tespit edilmiştir. Gelecekte modelin başarımının, farklı hiper parametreler ve mimariler ile de ölçülmesi hedeflenmektedir.

KAYNAKLAR

[1] Sabour, S., Frosst, N. ve Hinton, G.E., "Dynamic Routing Between Capsules", arXiv preprint arXiv:1710.09829, 2017.

[2] Xi, E., Bing, S. ve Jin, Y., "Capsule Network Performance on Complex Data". arXiv preprint arXiv:1712.03480, 2017.

[3] Hinton, G. E., Krizhevsky A. ve Wang, S. D. "Transforming Auto-encoders." International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2011.

[4] Hinton, G. E. "A Parallel Computation That Assigns Canonical Object-based Frames of Reference.", Proceedings of the 7th International Joint Conference on Artificial Intelligence-Volume 2., 1981.

[5] LeCun, Y., Cortes, C. ve Burges, JB "The MNIST Database of Handwritten Digits", 1998.

[6] Krizhevsky, A. ve Hinton, G.E. "Learning Multiple Layers of Features from Tiny Images.", 2009.

[7] Mavi, A. ve Dikle, Z., "Sign Language Digits Dataset", Ayrancı Anadolu Lisesi, Ankara, Türkiye, 2017, (https://github.com/ardamavi/Sign-Language-Digits-Dataset), (Erişim Tarihi: 19.01.2018)

[8] Fukushima, K. N. "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.," Biol. Cybern., vol. 36, no. 4, pp. 193–202, 1980.

[9] CapsNet-Tensorflow, Kapsül Ağı ve Klasik Sinir Ağı Karşılaştırması, (https://github.com/naturomics/CapsNet-Tensorflow), (Erişim Tarihi: 24.01.2017)

[10] Hubel D.H. ve Wiesel T.N. "Receptive fields and functional architecture of monkey striate cortex.", The Journal of physiology, 195(1):215-43, 1968.

[11] Qiao K, Zhang C, Wang L, Yan B, Chen J, Zeng L, Tong L. "Accurate reconstruction of image stimuli from human fMRI based on the decoding model with capsule network architecture", arXiv preprint arXiv:1801.00602. 2018.

[12] CapsNet-Keras, (https://github.com/XifengGuo/CapsNet-Keras?files=1), (Erişim Tarihi: 24.01.2017)

978-1-5386-1501-0/18/$31.00 ©2018 IEEE