

Uyarlanır Yerel Bağlı Nöron Modelinin İncelemesi

Araştırma Makalesi/Research Article

 Faik Boray TEK

Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Işık Üniversitesi, 34980 İstanbul, Türkiye
boray.tek@isikun.edu.tr

(Geliş/Received:24.05.2019; Kabul/Accepted:03.09.2019)

DOI: 10.17671/gazibtd.569827

Özet— Bu çalışmada uyarlanır yerel bağlı (odaklanan) nöron modelinin bir incelemesi sunulmuştur. Öncelikle bu modelin varolan diğer nöron modelleri ile ilişkisi incelenmiştir. Daha sonra modelin ileri beslemede çalışması ve geriye yayılım ile eğitilmesi tartışılmıştır. Modelin çalışma prensipleri sentetik sınıflandırma veri kümeleri üzerinde deneylerle gösterilmiştir. Son olarak, basit ve evrişimli ağların saklı katmanlarında odaklı nöronlar kullanılması halinde tam bağlı nöronlara göre daha iyi bir performans elde edilebileceği MNIST, CIFAR10, FASHION gibi popüler imge tanıma veri kümelerinde karşılaştırmalı olarak gösterilmiştir.

Anahtar Kelimeler— yapay sinir ağı, nöron, uyarlanır yerel bağlı nöron, odaklanan nöron

A Study on Adaptive Locally Connected Neuron Model

Abstract— The manuscript presents a detailed study of adaptive local connected (focusing) neuron model. Our analysis starts with the model's relation to other neuron models. Then we describe the feed-forward operation and its training with backpropagation gradient descent algorithm. The operation principles of the model were demonstrated with synthetically sampled data sets. Finally, the comparative experiments on popular image recognition datasets such as MNIST, CIFAR10, and FASHION show that using focusing neuron layers can improve the classification performance in some data sets.

Keywords— artificial neural networks, neuron, adaptive locally connected neuron, focusing neuron

1. GİRİŞ (INTRODUCTION)

Günümüzde otomatik yüz tanıma [1], konuşma tanıma [2] doğal dil işleme [3], Go oyunu [4] gibi birçok karmaşık problem derin yapay sinir ağları ile çözülebilmektedir [5]. Ülkemizde de uygulama örnekleri yaygınlaşmaktadır [46,47]. Bu gibi problemlere başarı ile uygulanan yapay sinir ağları mühendisler tarafından sezgisel ve karmaşık yöntemler kullanılarak tasarlanır.

Genelde, kullanılan ağ topolojisi (yapısı) öğrenme öncesinde planlanır, kurulur ve öğrenme sırasında değişmez. Topolojinin sabit olması ağ kurulduktan sonra ve eğitim sırasında bilgi akışının değişmeyeceği anlamına gelir. Ağ yapısını eğitimin bir parçası yapan çalışmalar bulunmaktadır. Bunlardan bazıları eğitim

sırasında ağa nöron ekleme veya çıkarma (budama) üzerine kuruludur [6,7,8]. Bazıları birden fazla ağ topolojisini eğiterek doğrulama kümesindeki performanslarına göre en iyi ağ yapısını seçmeyi hedefler [9,10,11,12]. Makalemizde ise ağ topolojisini eğitime dahil eden uyarlanır yerel bağlantılar kurabilen yeni bir nöron modeli tartışılmaktadır. Yerel bağlantı alanının hem açıklığının, hem de konumunun değişebilir olması dolayısıyla yeni nöron modeline odaklanan nöron adı da verilmiştir [13, 45].

Yeni nöron modeli giriş sinyallerin uzamsal boyutu olduğunu varsayar. Bir başka deyişle giriş sinyallerinin konumları bulunur. Bu konum uzayında değişebilen ve türevlenebilen bir odak işlevi sayesinde nöron, alıcı alanını (ing. receptive field) eğitebilir. Bu sayede

uyarlanır bir yerel bağlantı yeteneği sunar. Odaklanan nöronlardan oluşan bir katman (veya ağ) probleme özgü bağlantı yapıları kurabilir.

Bu makalede iki önemli soruya yanıt verilmeye çalışılmıştır: 1) Odaklanan nöronlar anlamlı girdilere yönelecek şekilde yerel bağlantılarını değiştirebilir mi? 2) Bu yetenek, ağın sınıflandırma ve genelleme başarısını (tam bağlı nörona göre) iyileştirebilir mi?

İlerleyen bölümlerde biyolojik nöron temellerinden başlanarak, tam bağlı yapay sinir ağı ve eğitimi yerel bağlar perspektifinden incelenecek ve önerilen modelin tartışılacağı yöntem kısmı için bir temel oluşturulacaktır. Daha sonra uyarlanır yerel bağlı (odaklanan) nöron modeli detaylı bir şekilde incelenecek ve sınanacaktır.

2. İLGİLİ ÇALIŞMALAR (BACKGROUND)

Beyin milyarlarca sinir hücresinden (nöron) oluşan yapısı ile bütünlük bir yapıdan ziyade birçok işlevsel bölgeye ayrılmış bir meta-organdır [14,15]. Örneğin sol yarımkürede yer alan ön lob konuşma, yazı, planlama, duygular gibi işlevleri yerine getirirken, oksipital lobda görsel bilgi işleme ve okuma gibi işlevler, sereblumda ise istemli hareket koordinasyonu ve denge işlevi yerine getirilir. İnsan için çok zor olmayan bisiklet sürme gibi bir işlevi ele alırsak, kişi aynı anda gerekli el ve ayak hareketlerini yaparak dengeli bir şekilde ilerleyebilir, diğer araçların seslerini kuşların sesinden ayırt edebilir, yolda yeralan çukurları tanır ve kaçınmak için planlama yapabilir. Bütün bu işlevleri aynı anda yapabilmesinin nedeni beynin farklı bölgelerinde, farklı sensörlerden (sinir hücrelerinden) gelen bilgileri paralel bir şekilde işleyebilmesi ve bölümler arasında bilgi taşıyabilmesidir [16].

Bunun yanında biyolojik nöronlar genelde tek tip gibi gösterilse de hem işlevsel hem de yapısal farklılıklar gösterir. Bazı nöronlar diğerlerine göre daha çok alıcı dendrite sahipken (multipolar), bazı nöronlar bölümler ve konumlar arasındaki bilgi taşıma görevini üstlendiğinden iki kutuplu ve daha uzun aksonlara sahiptir (bipolar).

Beynin sürekli gelişerek binlerce işlevi yapabilmeyi öğrenmesini sağlayan sistem ise nöroplastiktir. Beyin yeni problemlerin çözümü için çalıştıkça yeni yerel (lokal) bağlantılar kurar, gerek görülmeyen bazı bağlantıların zayıflamasına izin verir [17]. Aksonların doğru hedef hücreleri bulmak için mekanik ve biyokimyasal ipuçlarını/gradyanları izlediklerini gösteren çalışmalar mevcuttur [18]. Makalemize konu olan uyarlanır yerel bağlı (odaklanan) nöron, biyolojik nöronun bu özelliğine öykünerek bağlantı kuracağı nöronları eğitim sırasında belirler.

Çok gelişmiş olsalar dahi yapay sinir ağları genellikle tek bir problemi çözebilecek bir işlevi modeller. Problemi çözeceği düşünülen ileri beslemeli bir ağ kurulur ve

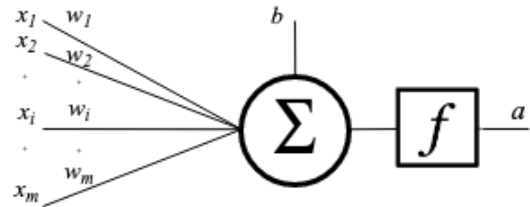
eğitilecek parametreler belirlenir. Eğitim sonunda, girdi ve parametre değerleri ile alınan çıktı ile beklenen çıktı arasındaki farkı (hatayı) minimize edecek parametre değerleri çözülmeye çalışılır. Belirli problemler için yapı bir kez kurulur, eğitilir ve sonra kullanılır. Transfer öğrenme sayesinde, derin yapay sinir ağı mimarilerinde eğitilmiş olan bağlantı ağırlıkları farklı problemlere transfer edilebilmekte ve bir miktar daha eğitim ile başarılı sonuçlar alınabilmektedir [19]. Ağ modelleri farklı problemlere uygulanabilse de ağların koordinasyonu ile beyindeki benzer şekilde öğrenen yöntemler çok yaygın değildir. Ancak ağ yapısını özyineleyen şekilde kurgulayan, evrimsel metodlarla veya pekiştirmeli öğrenme ile öğrenen çalışmalar bulunmaktadır [20,21,22]. Bu makalenin konusu ise ağın yapısını eğitimin içerisine doğrudan dahil eden çekirdek bir nöron modelidir.

2.1. Tam Bağlı Nöron (Fully Connected Neuron)

Tam bağlı bir nöron girdi uzayında bütün girdilere bağlıdır [23]. Şekil 1'de gösterildiği gibi nöronun parametreleri (w_i), giriş sinyallerinin nasıl ağırlıklandırılacağını belirler. Nöronun eğitilebilir parametrelerini ve tüm ağın hatasının yerel çözüm alanını oluşturur. Nöron diğer m nörondan kendisine ulaştırılan girdileri ağırlıklarla çarpıp, toplar ve bir transfer işlevinden (f) geçirerek çıktısı a 'yı hesaplar (1).

$$a = f\left(\sum_i^m w_i x_i + b\right) \quad (1)$$

Şekil 1'de gösterilen nöronun sınıflandırma veya yordama için kullanılan bir ileri besleme görevi vardır.



Şekil 1. Tam bağlı nöron: girdilerini, ağırlıklarıyla çarpıp toplar ve bir aktivasyon transfer işlevinden geçirerek çıktısını hesaplar (Fully connected neuron model)

2.2. Hatanın Geriye Yayılımı ile Eğitim (Training with Error Backpropagation)

Nöronun bağlantı katsayıları (w_i) geriye yayılım algoritması kullanılarak hatayı azaltacak yönde güncellenerek öğrenme gerçekleştirilir [23]. Buradaki nöronun yapay sinir ağının çıkış katmanındaki n nörondan biri olduğunu kabul edelim. Dolayısıyla a_j çıkış değerini, t_j ise hedef değeri gösterebilir. Toplam hata E , a_j ile t_j 'nin kare hata farkının bütün çıktı nöronları üzerinde ortalamasıdır (2). Hata işlevi ağın çıktı değişkenine göre türevi tanımlı herhangi başka bir işlev olabilir [23].

$$E = \frac{1}{2} \sum_{j=1}^n (t_j - a_j) \quad (2)$$

Geriye yayılım ile eğitim sırasında nöronun hataya katkısı zincir kuralı ile belirlenir (3).

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial f} \frac{\partial f}{\partial w_i} \quad (3)$$

Hataya katkısı belirlenen ağırlık gradyan yönünün tersinde η ölçeğinde bir adım atarak güncellenir:

$$\widehat{w}_i = w_i - \eta \frac{\partial E}{\partial w_i} \quad (4.1)$$

$$\widehat{w}_i = w_i - \eta \frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial f} x_i \quad (4.2)$$

Tam bağlı nöron modelinde (1) bağlantı ağırlıkları hem nöronların birbirine bağlı olup olmadığını hem de bağlantı şiddetini temsil eder. Modelde bir veya daha fazla bağlantı ağırlığının sıfır olmasına bir engel yoktur. Ancak pratikte geriye yayılım algoritması ve kullanılan hata işlevi nedeniyle eğitim sırasında bu durum oluşmaz.

Bu durumu gradyan iniş güncelleme ifadesi (4.2) üzerinden inceleyebiliriz. Aynı katmanda yer alan iki farklı ağırlık w_1 ve w_2 olsun. Bu ağırlıklardan bir tanesi anlamlı bir girdiye, diğeri ise sabit anlamsız bir girdiye bağlı olsun ($x_2 = c$). Eğitimin başında anlamsız bir girdiye bağlı olduğu için $w_2 = 0$ atayalım. Bir sonraki iterasyonda w_2 tekrar sıfır olmayan bir değere güncellenecektir. Bunun nedeni güncelleme ifadesinde sağ tarafta yer alan iki türevin ($\frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial f}$) w_1 ve w_2 için ortak olmasıdır. Bir başka deyişle, w_1 güncellendiğinde, w_2 'de güncellenecektir. Sabit giriş x_2 sıfır değilse, ağırlık w_2 'nin sıfır kalabilmesi için tek bir durum söz konusudur: eğitimin sonuna gelmiş ve hata sıfırlanmıştır $\partial E / \partial a_j = 0$. Bu durumda her iki ağırlık güncellenmez.

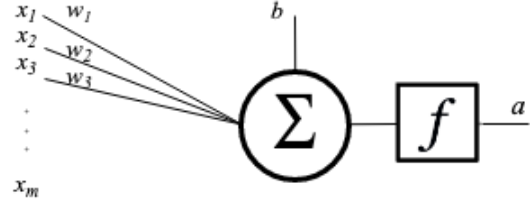
Eğitim süresince iterasyonlarla w_2 sıfıra yaklaşabilir mi? Ortak gradyan $\partial E / \partial a_j$ 'a ek olarak w_2 nin güncellenmesini etkileyen faktör x_2 , eğitim süresince sabit olacağından (rastgele de olabilir) bu durum da gerçekleşemez.

Ağırlıkların toplam büyüklüğü hataya eklenirse (örn. L_1 , L_2 normları ile) küçültme yönünde bir gradyan oluşacaktır (5). Bu durumda bütün ağırlıkların küçülmesi ve sıfıra yaklaşması teşvik edilir. Ancak bu durumda α dikkatli seçilmelidir. Çünkü amaçlanan hatayı (2) azaltmak için bir bağlantı ağırlığını artırmak gerekirken, doğrultma yüzünden azaltmak gerekiyor olabilir.

$$E = \frac{1}{2} \sum_{j=1}^n (t_j - a_j) + \alpha L_n(\mathbf{w}) \quad (5)$$

2.3. Yerel Bağlı Nöron (Locally Connected Neuron)

Yerel bağlı nöronun tam bağlı nörondan farkı girdi uzayının sadece sınırlı bir bölgesine bağlantı kurulmasıdır [24]. Bu özellik ile uzamsal sabitlik gösteren problemlerde farklı istatistiksel özellikler çıkartılmasını kolaylaştırır [25]. Örneğin yüz tanıma probleminde gözler ve ağız arasında uzamsal bir ilişki bulunur [26].



Şekil 2. Yerel bağlantılı nöron modeli.
(Locally connected neuron model)

Tam bağlı nöron ile ileri ve geri besleme çalışma prensipleri aynıdır, bağlantı kurulan nöron sayısı (s) geri katmanda bulunan bütün nöronlar olarak belirlenirse tam bağlı nörona denktir. Şekil 2'de gösterilen örnek, denklem (6)'da ifade edilen nöron sadece ilk üç (s) girdiye bağlıdır. Genelde tüm girdi uzayı, eş sayıda ve örtüşen bölgelere ayrılarak bölünür ve/veya nöronların bağlantı yapacağı nöronlar sabitlenir. Bu eğitimden önce ağ kurulurken yapılır ve eğitim boyunca bağlanan nöronlar ve bağlanan nöron sayısı değişmez. Eğitim tam bağlı nöronlarla aynıdır, bağlı olan ağırlıklar hataya katkıları uyarınca güncellenir.

$$a = f \left(\sum_{i=1}^{s \leq m} w_i x_i + b \right) \quad (6)$$

Üstteki yerel bağlı nöronun farklı olarak uyarılar yerel bağlantılı model bağlantı konumunu ve açıklığı değiştirebilir ve geriye yayılım algoritması ile eğitilebilir.

Tam bağlı nöronun ve yerel bağlı nöronun bağlantı ağırlıkları kendi aralarında bağımsızdır. Ancak, evrimsel nöronlarda bağımsız ağırlıklar yerine paylaşımlı ağırlıklar kullanılır. Başka bir deyişle girdi uzayının farklı bölgelerinden bilgi toplayan nöronların bağlantı ağırlıkları ortak/paylaşımlıdır [5,27,28]. Bu sayede oluşan evrişim çekirdekleri (süzgeçler) girdi uzayının farklı bölgelerine aynı işlemi uygular. Bu sayede girdi uzayında ötelemeden bağımsız bir temsil elde edilir. Evrimsel çok katmanlı mimarilerin biyolojik görü sisteminde iç temporalde bulunan yapıyı temsil ettiği gösterilmiştir [29].

2.4. Yerel Almaç Alanları (Local Receptive Fields)

Rosenblatt'ın perseptron modelinden [30] bu yana farklı nöron modelleri tartışılırken yerel almaç alanları da söz konusu olmuştur. Bir grup çalışma, nöronun girdi değer uzayının bir kısmında seçici olarak aktif olmasını yerel almaç alanı olarak tanımlar [31]. En tanınmış örneklerden biri radyal baz işlevli (ing radial basis function, RBF) nöronlardan oluşan yapay sinir ağıdır [32].

RBF nöron modeli odaklanan nöron modeline yakın görünse de baz işlevlerini girdi (değer) uzayını kümelendirerek, nöronları örneklere özelleştirmek için kullanılır. Odaklanan nörona ise (yöntem kısmında anlatılacağı gibi) Gauss işlevi nöronun bağlantı konumunu (yerelliğini) belirlemek için kullanılmıştır. Kendini organize eden harita (ing. self organizing map, SOM) ağlarında giriş değer uzayı iki boyutlu bir latıs üzerinde yerleşmiş nöronlar üzerine izdüşümlenir [33]. Dolayısıyla SOM nöronlarının kendi uzamsal boyutları bulunur, ancak girişin uzamsal boyutu yok sayılır.

2.5 Uyarlanır Yerel Bağlı Modeller (Adaptive Local Connected Models)

Bu kısımda bahsedeceğimiz çalışmalar odaklanan nöron da olduğu gibi yerel almaç alanını topolojik (konum) uzayında tanımlar. Uyarlanır yerel bağlantının biyolojik beyinde karşılığı nöroplastiktir. Örneğin nöroplastikliği Hebb kuralı ile uygulama denemeleri olmuştur [35,36].

Öte yandan görsel dikkatin (ing. visual attention) modellenmesi probleminde konuma duyarlı ağ tasarımları sıkça görülen bir yaklaşımdır [37-42]. Bu tür yöntemlerde girdi uzayının farklı konumlarına dinamik olarak odaklanabilen özyineleyen bir ağ yapısı kurulur. Bir giriş imgesi n yinelemede her seferinde farklı yerel bir bölgeye odaklanarak işlenir.

Cheung [41] çalışmasında görsel dikkat için Gauss çekirdek fonksiyonları kullanılarak bir retinal gözetiş modeli sunulmuştur. Bu yöntemde odaklanan nörona olduğu gibi Gauss çekirdekleri giriş uzayının farklı bölgelerinden sinyal almak için kullanılmıştır.

3. YÖNTEM (METHOD)

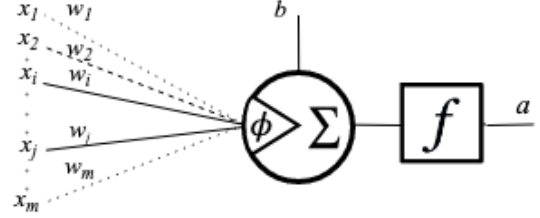
3.1 Uyarlanır Yerel Bağlı Nöron: Odaklanan Nöron (Adaptive Locally Connected Neuron: Focusing Neuron)

Uyarlanır yerel bağlı nörona eğitilebilir bir odak işlevi bulunur. Bu odak sayesinde nöron girdi uzayının sınırlı bir bölgesiyle bağlantı kurar. Bir bölgeden gelen sinyaller toplanırken diğer bölgelerden gelen sinyaller sönmüldürülür. Odak eğitilebilir, yani çıktı hatasını azaltacak bir yönde değişebilir. Nöron odaklandığı bölgenin konumunu değiştirebildiği gibi bölgenin genişliği de (açıklığı) eğitim sırasında değiştirebilir. Dolayısıyla nöron gerekirse tam bağlı gibi şekillenip geniş bir bölgeye veya bir kaç nörondan oluşan dar bir bölgeye odaklanabilir. Bir boyutlu bir girdi uzayında odaklanan nöron modeli Şekil 3'de gösterilmiştir.

Girdi uzayında yer alan sinyal kaynağı i indisi ile, bu kaynağın sürekli bir uzayda konumunu belirten işlev $\tau(i) \in \mathcal{R}$ ile gösterildiğinde, odaklanan nöron işlevi aşağıdaki gibi ifade edilebilir (8):

$$a = f\left(\sum_i^m w_i x_i \varphi(\tau(i), \theta) + b\right) \quad (8)$$

$\varphi(\tau(i), \theta)$ girdinin konumu ve θ parametre vektörü değerine göre katsayı üreten deterministik bir işlevdir. Bu katsayılar her bağlantıya ek bir ağırlık atamayı andırırsa da, birbirlerine ve girdi konum uzayına bağımlı olduklarından oldukça farklı bir işlev görmektedir. Bunun yanında bağlantı ağırlıklarıyla (w_i) uyum içerisinde eğitilebilirler. Odak bölgesi içerisine düşen sinyalin net girdiye hangi kuvvet ve işaretle ekleneceğini belirlemek w_i nin görevidir.



Şekil 3. Uyarlanır yerel bağlı (odaklanan) nöron modeli. (Adaptive locally connected (focusing) neuron model)

3.1. Odak İşlevi Örneği (An Example Focus Function)

Bu çalışma kapsamında i indisi girdinin konumu $\tau(i)$, normalize edilmiş konum olarak alınabilir: $\tau(i) = i/m$. Bir katmanda yer alan j indeksli odaklanan nöron için işlev $\varphi(i/m, \theta_j)$ şeklinde yazılabilir. Burada i parametresi bağlanılan girdinin kendi katmanındaki indisini, m katmandaki toplam nöron (girdi) sayısını, θ_j ise j indisli odaklanan nöronun kontrol parametrelerini gösterir. $\varphi(i/m, \theta_j)$ işlevinin türevinin $\partial\varphi/\partial\theta_j$ tanımlı olması değerlerinin geriye yayılım sırasında gradyan alçılması algoritmasıyla öğrenilmesine (9) izin verecektir.

$$\nabla\theta_j = \frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial f_j} \frac{\partial f_j}{\partial \varphi} \frac{\partial \varphi}{\partial \theta_j} \quad (9)$$

$$\hat{\theta}_j := \theta_j - \eta \nabla\theta_j$$

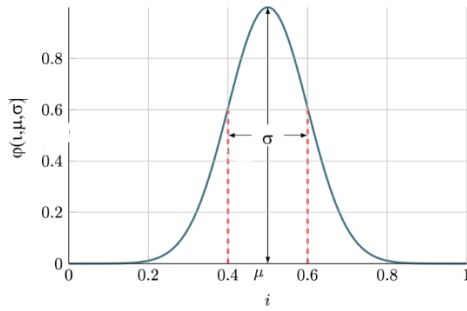
Odak kontrolü φ işlevi için birden fazla aday bulunmasına rağmen Gauss formu sürekli türevlenebilir olması dolayısıyla gradyan alçılması yöntemiyle eğitmek için idealdir. Şekil 4'te Gauss formu bir boyutlu odak işlevi gösterilmektedir.

Bu işlevde parametreleri $\theta = \{\mu, \sigma\}$ sırasıyla merkezi ve açıklığı ifade eder (8).

$$\varphi(i/m, \mu, \sigma) = s e^{-\frac{(i-\mu)^2}{2\sigma^2}} \quad (10)$$

Merkezi gösteren μ parametresi ile odağın konumu; açıklığı gösteren σ parametresi ile de genişliği değiştirilebilir. Genellikle normalize etmek için kullanılan s parametresi ise nöronun toplam girdisini, aynı büyüklükteki bir girdi uzayından sinyal toplayan tam

bağlı bir nörona, denkleştirmek için kullanılır ve eğitilmesi gerekli değildir.



Şekil 4. Odaklayan işlev örneği: konumu i ile belirtilen nöronlardan μ değerine yakın olan nöronlar güçlü bağlantılar kurabilirler; uzaktaki girdiler sönmüldürülür. İndis $[0,1]$ arasında normalize edilmiştir.

(An example focus function in Gauss form. Focus allows inputs positioned close to its center and eliminates those which are farther.)

Gradyan alçalmasına dayalı parametre güncelleme kuralı $d\varphi/d\mu$ ve $d\varphi/d\sigma$ kısmi türevleri kullanılarak merkez ve açıklık için ayrı ayrı uygulanır (11). Ağırlıkların w_{ij} güncellenmesi sırasında φ ek bir çarpan olarak (11.8) gelmektedir.

$$\frac{\partial E}{\partial \mu_j} = \frac{\partial E}{\partial \varphi} \frac{\partial \varphi}{\partial \mu_j} \quad (11.1)$$

$$\frac{\partial \varphi}{\partial \mu_j} = \frac{i/m - \mu}{\sigma_j^2} \varphi(i/m, \mu_j, \sigma_j) \quad (11.2)$$

$$\hat{\mu}_j = \mu_j - \eta \frac{\partial E}{\partial \mu_j} \quad (11.3)$$

$$\frac{\partial E}{\partial \sigma_j} = -\eta \frac{\partial E}{\partial \varphi} \frac{\partial \varphi}{\partial \sigma_j} \quad (11.4)$$

$$\frac{\partial \varphi}{\partial \sigma_j} = -\frac{\left(\frac{i}{m} - \mu\right)^2}{\sigma_j^3} \varphi(i/m, \mu_j, \sigma_j) \quad (11.5)$$

$$\hat{\sigma}_j = \sigma_j - \eta \frac{\partial E}{\partial \sigma_j} \quad (11.6)$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial f} \frac{\partial f}{\partial w_{ij}} \quad (11.7)$$

$$\frac{\partial f}{\partial w_{ij}} = x_i \varphi(i/m, \mu_j, \sigma_j) \quad (11.8)$$

$$\hat{w}_{ij} = w_{ij} - \eta \frac{\partial E}{\partial w_{ij}} \quad (11.9)$$

4. DENEYLER (EXPERIMENTS)

Öncelikle, odaklanan nöronun çalışma prensibini göstermek için sentetik örneklerle çalışılmıştır. İkincil olarak, popüler imge tanıma veri kümelerinde öğrenme ve genelleme performansları sınanmıştır. Deneyler Python programlama dilinde, numpy, Theano [43] ve Lasagne kütüphaneleri kullanılarak kodlanmıştır. Deneylerin tekrarlanabilmesi için kodlar paylaşılmıştır [44].

4.1. Sentetik Veri (Synthetic Data)

Bu deneyde amaç, girdi uzayına bilgi içermeyen rastgele girdiler eklediğimizde odaklanan katmanlarda yer alan nöronların bilgi içermeyen girdilerden uzaklaşım uzaklaşamayacağını gözlemlemektir.

İlk deney için normal dağılımlı izotropik 30 boyutlu bir uzaydan rastgele örnekleme yöntemi ile sentetik veri oluşturuldu. Örneklemede dört ayrı farklı öbekten (sınıftan) toplam 5000 nokta oluşturuldu. Veri kümesindeki her bir örneği $\mathbf{x} = \{x_0, x_1, \dots, x_n\}$ şeklinde ifade edersek, ilk deneyde vektörün sağına ve soluna bilgi içermeyen rastgele değerler eklenerek $\mathbf{x}' = \{r_0, \dots, \mathbf{x}, \dots, r_n\}$ yeni bir veri kümesi oluşturuldu; sonraki deneyde ise sadece soluna rastgele değerler eklenerek $\mathbf{x}' = \{r_0, \dots, r_n, \mathbf{x}\}$ yeni bir veri kümesi oluşturuldu. Böylelikle her bir örneğin boyutu 60'a ulaşmıştır. Veri kümeleri, rastlantısal olarak %70 eğitim ve %30 doğrulama kümesi olarak iki kısma ayrılmıştır.

Veri kümesini öğrenmek üzere odaklı saklı katman (OK) ve kontrol için tam bağlı (TB) saklı katmandan oluşan iki ayrı basit ağ yapısı oluşturulmuştur:

Tam bağlı katmanlardan oluşan ağ:

Girdi (60)-> **TB** (10 nöron) -**TB** (4 nöron, çıktı)

Odaklanan nöron katmanlı ağ:

Girdi (60)-> **OK** (10 nöron) -**TB** (4 nöron, çıktı)

Başlangıç odak konum (μ) değerleri girdi katmanı konumlarına eşit aralıklarla dağıtılmıştır ([0.2-0.8]). Odak başlangıç açıklık değeri $\sigma=0.15$ olarak belirlenmiştir. Eğitim 512 mini-yığın örnek sayısı ve stokastik gradyan alçalma algoritması ile gerçekleştirilmiştir. Güncellemelere ek bir koşul konularak merkezin μ [0-1.0] aralığında kalması, açıklığın σ [0.01-0.5] aralığında kalması sağlanmıştır. Bunun nedeni, ilk bir kaç güncellemede gelen yüksek gradyan değerlerinin nöronu işlevsiz bir duruma sokmasını engellemektir. Öğrenme katsayıları ağırlıklar için 0.1, μ ve σ için ise 0.01 ve 0.001 olarak belirlenmiştir.

İlk deneyde, bu parametrelerin eğitimden önceki ilk durumu Şekil.5a'da gösterilmiştir. Eğitim sonrasındaki durumu Şekil 5b'de incelendiğinde odakların genel olarak merkeze toplandığı, bazı odakların genişlediği bazılarının daraldığı gözlenmektedir. Şekil 5c ve 5d'de odak parametrelerinin eğitim süresince değişimi gözlenmektedir. Bu grafikler incelendiğinde ilk epoklarda

hızla başlayan konumlanma eğiliminin epoklar ilerledikçe yavaşladığı ve görece durağanlaştığı görülebilir. Açıklık değişim grafiklerinde ise çoğunluk odağın yavaşlayan bir açılma, büyüme eğiliminde oldukları gözlenmektedir. Sonuç olarak odakların anlamlı bilginin olduğu merkez konumlara doğru kaydığı ve ilk durumlarına göre genişledikleri gözlenmiştir.

Bu sonuçların anlamlı veriye yönlenmeyi mi, yoksa sadece merkeze toplanma eğilimini mi gösterdiğini anlamak için ikinci kurguda örnek uzayının sol tarafına rastgele gürültü değerleri eklendi. Yönelimleri daha iyi gözlemlemek için başlangıç odak merkezleri girdi uzayının merkezinden başlatıldı. Beklenti, odakların girdi uzayının sağ tarafına yönelmesidir. Ağ ve eğitim parametreleri bir öncekiyle aynıdır.

Şekil 6a ve 6b'de odakların ilk ve eğitim sonrası durumları görülmektedir. Burada üçü hariç diğer odakların sağa doğru kaydığı gözlenmiştir. Şekil 6c ve 6d'de gösterilen eğitim sırasında değişim grafikleri incelendiğinde odakların ilk bir kaç epokta hızla değiştiği, daha sonra yavaşlayarak merkez konumlarının yakınsadığı gözlenmektedir. Odak açıklıklarının genelde yavaşlayan bir artma eğiliminde olduğu gözlenebilir.

İlk iki deneyde odakların rastgele veri konumlarından uzaklaşarak anlamlı veriye yöneldiği görülmüştür. Bu yeteneğin sınıflandırma ve genelleme başarısı için bir avantaj sağlayıp sağlamadığını ortaya koymak için bir üçüncü deney kurgulandı. Bu deneyde parametreleri aşağıdaki Tablo 1.'den rastgele seçilmiş 250 farklı durum için 3 kez farklı rastgele başlangıç ile veri kümesi sentezleyen, ağ kuran, eğiten ve ayırık bir kümede test eden bir kod oluşturuldu. Her bir durum için odaklanan katmanlı ağ (OA) ve tam bağlı ağ (TBA) doğrulama kümesinde doğru sınıflandırma yüzdesi kaydedildi. Şekil 7a her bir karşılaştırmayı $x=OA$ başarısı, $y=TBA$ başarısı şeklinde bir nokta olarak göstermektedir. Gürültü eklenmiş veri kümelerinde ağırlıklı olarak odaklı ağ daha yüksek doğruluklu sınıflandırma yapabilmektedir. Temiz verilerde ise tam bağlı ağa göre bir avantaj gözlenmemektedir.

Şekil 7b'de eğitim süreleri karşılaştırmalı çizilmiştir. Odaklı ağ ve tam bağlı ağ epok eğitim süreleri doğrusala yakın bir ilişkidir. Şekil 7c OA'nın ortalamada TBA'ya göre 1.3 kat daha yavaş eğitildiğini göstermektedir. Bu süre farkı odak merkez ve açıklık parametrelerinin gradyanlarının hesaplanmasının getirdiği ek maliyettir. Odak saklı nöron başına ek iki parametre getirmektedir. Girdi sayısından bağımsız olduğundan örnek sayısı ile ilişkisinin doğrusal olması beklenen bir durumdur.

4.2. MNIST Veri Kümesi (MNIST Data Set)

Bu kısımda sentetik veriler ile yapılan ilk deney MNIST karakter tanıma veri kümesi ile tekrarlanmıştır. Bu küme 28x28 boyutunda, gri seviye 60000 adet el yazısı karakter imgesinden oluşmaktadır. 50000 imge eğitim, 10000

imge doğrulama, 10000 imge ise sınama için ayrılmaktadır. Bu amaçla önceki gibi yine iki katmanlı bir ağ kurularak (giriş, saklı, çıkış) odakların eğitim ile nasıl farklılaştığı gözlemlenmiştir.

Tablo 1. Sentetik veri, ağ ve eğitim parametre kümeleri.
(Synthetic data set, network and training parameter sets)

| | |
|---------------------------------------|-------------------------------|
| Girdi Örnek sayısı | $N=\{100, 200, 2000, 10000\}$ |
| Girdi örnek boyutu | $nF=\{4, 40, 100\}$ |
| Gürültü eklenti boyutu | $nF*\{0.0, 0.5, 1.0\}$ |
| Gürültü pozisyonu | {Yanlar, Sol} |
| Öbek sayısı | {1, 2, 4} |
| Sınıf sayısı | {2, 4, 8} |
| Saklı nöron sayısı | {4, 40, 100} |
| Eğitim mini yığın örnek sayısı | $N*\{0.1, 0.05\}$ |
| Eğitim hızı | {0.01, 0.001} |
| L2 Cezası (ing. Penalty) | {Doğru, Yanlış} |

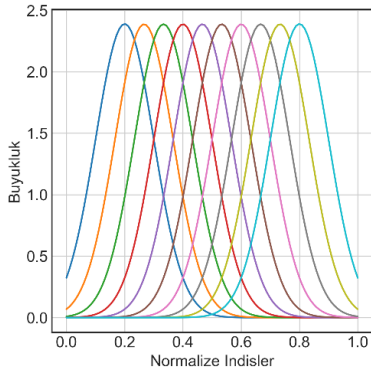
Şekil 8a'da, girdi uzayının orta konumundan başlatılan, ilk açıklıkları 0.1 olan, odakların 250 epok eğitim sonundaki durumları gözlenmektedir. Şekil. 8b'de merkezlerin değişiminin, sentetik verilerle benzer olduğu gözlenmektedir. Burada odaklar başlangıçta hızla yayılarak bütün girdi uzayına yayıldıklarını gözlenmiştir. Ancak bu durum odak merkez öğrenim hızına bağlıdır. Odakları, yayarak başlatmak genelde daha iyi sonuçlar vermesine rağmen eğitim sırasında yönelimlerinin açıkça gözlenebilmesi için burada merkezden başlatılmıştır.

Şekil 8c ve 8d'de eğitim ve sınıflandırma hata ve doğruluk yüzdeleri görülmektedir. Grafikler incelendiğinde odaklı ağın (OA) tam bağlı (TBA) ağa göre daha iyi sonuç verdiği görülebilir. Ancak bu sonuçlar tek bir rastgele düzen için ve sınırlı sayıda saklı nöron ile alınmıştır. Saklı nöron sayısı arttıkça veya farklı rastgele başlangıçlarda farklı sıralamalar oluştuğu da gözlenmiştir.

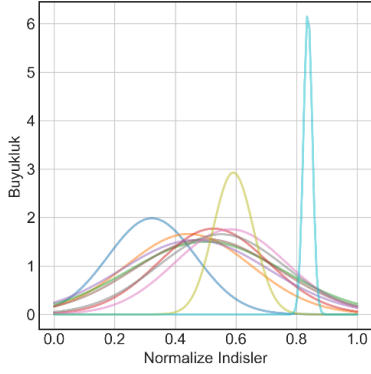
Daha iyi bir karşılaştırma ve güncel en iyi sonuçlara yakınsayan sonuçlar için daha gelişmiş ağ yapıları ile deneyler yapıldı: Bunlar, üç katmanlı basit ağlar (giriş, saklı, saklı, çıkış) ve evrimsel ağlardır (giriş, evrişim, havuz, evrişim, düşürme, saklı, çıkış).

Basit 3 katmanlı Ağ (BA): Giriş- **SK**(800)-DP(0.25)-**SK**(800)-DP(0.25)- Çıkış(10)

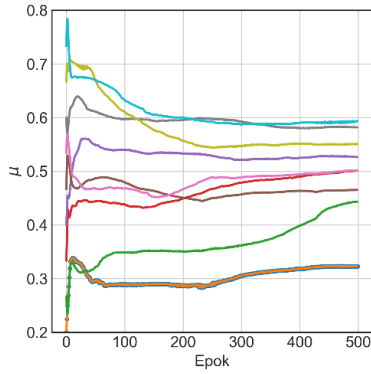
Evrimsel Ağ (EA) : Giriş- EV (32) – PL(2,2) - EV (32) – PL(2,2)- DP(0.25) - **SK** (250) – Çıkış(10)



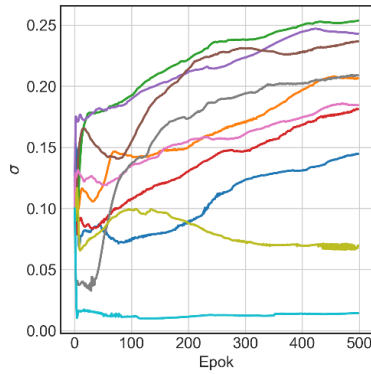
a)



b)



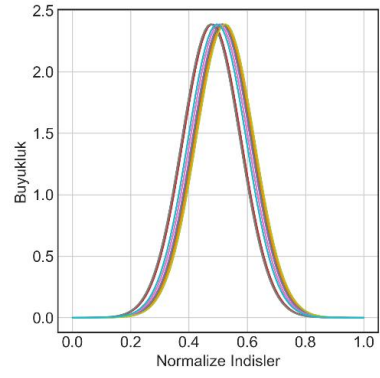
c)



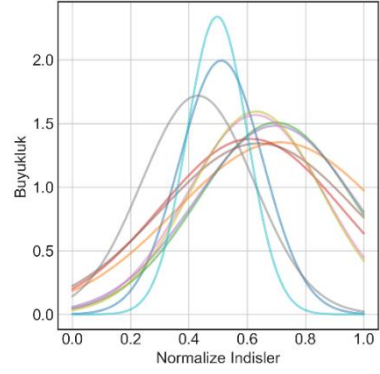
d)

Şekil 5. Sağ ve sol gürültülü sentetik veri kümesinde odakların eğitimle değişimi. a) ilk durum, b) eğitim sonunda durum, c) Odak merkezi değişimi, d) Odak açıklığı değişimi.

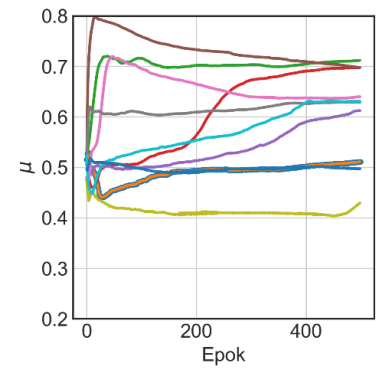
(Training of the focused neurons in left and right noise synthetic datasets.)



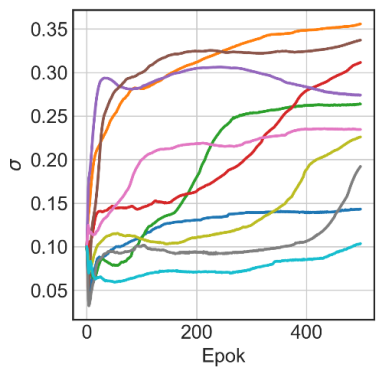
a)



b)



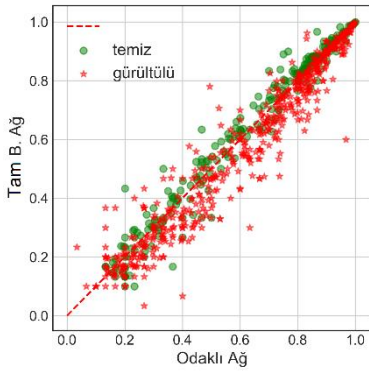
c)



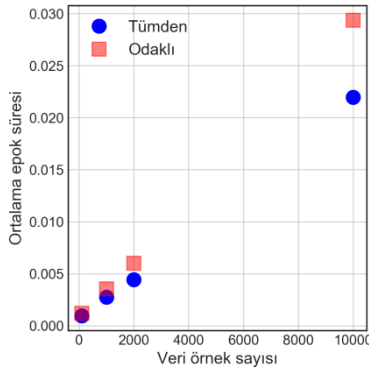
d)

Şekil 6. Sol gürültülü sentetik veri kümesinde odakların eğitimle değişimi. a) ilk durum, b) eğitim sonunda durum, c) Odak merkezi değişimi, d) Odak açıklığı değişimi.

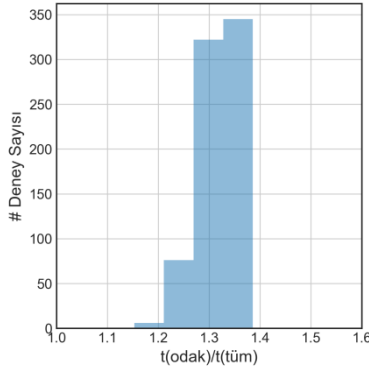
(Training of the focused neurons in left-added noise synthetic datasets.)



a)



b)



c)

Şekil 7. 750 farklı veri kümesi sınavında Tam bağlı ağ (TBA) ve odaklı ağ (OA) karşılaştırması. a) Doğrulama kümesinde doğru sınıflandırma yüzdelерinin saçılımı ($x=OA$, $y=TBA$). b) OA ve TBA epok süresinin veri sayısına göre değişimi. c) OA epok süresinin TBA epok süresine oranının dağılımı.

(Comparisons of a 2-layer fully connected layer network and focusing layer network in 750 synthetic datasets. a) Each point shows the classification accuracies in the validation data set, b) average training epoch time vs sample size, c) histogram of the epoch time ratios.)

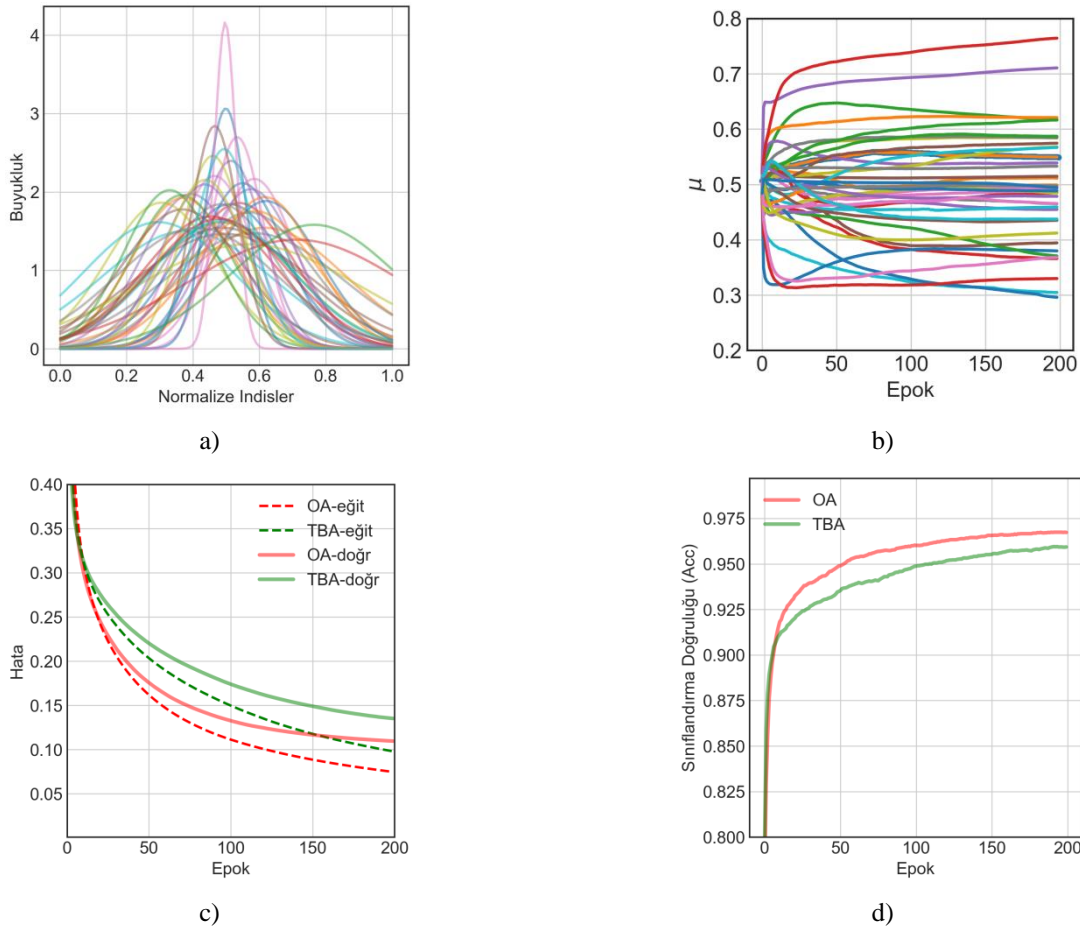
Her iki yapıda **SK**(n) (n nöronlu saklı katman) odaklanan ve tam bağlı nöronlarla kurularak karşılaştırıldı. Diğer kısaltmalar DP(p) (ing. Drop-out, p olasılıklı rastgele düşürme) katmanı, EV(n), n süzgeçten oluşan evrişim katmanını, PL(m,m) mxm lik havuzlama katmanını temsil etmektedir.

Deneyimizde kullanılan veri kümeleri şu şekildedir: 1) MNIST: karakter tanıma kümesi, 2) MNIST-CLUTTERED: MNIST örneklerinin 60x60 rastgele gürültülü arka planlara yerleştirilmiş hallerinden oluşan veri kümesi, 3) CIFAR-10: 32x32x3 renkli 60000 imgeden oluşan 10 sınıflı genel nesne tanıma kümesi, 4) FASHION: MNIST formatında hazırlanmış 10 sınıflı moda nesnelere (etek, pantolon, şort vs) veri kümesi. MNIST ve FASHION veri kümelerinde örnekler çerçeveye ortalanmıştır. Diğer veri kümelerinde nesne imge çerçevesinin farklı konumlarında bulunabilir.

Eğitimde kullanılan parametreler: öğrenme hızı (MNIST ve diğerleri: 0.01, FASHION: 0.0005), odak parametreleri için öğrenme hızları ($n_\mu=0.1$ ve $n_\sigma=0.01$) şeklindedir. Öğrenme hızı azaltımı: 100 (CIFAR10'da 10) epok sonunda önceki öğrenme hızları %90 azaltılmıştır. Momentum ile stokastik gradyan inişi kullanılmıştır. Her karşılaştırma rastgele 5 farklı kez başlatılmış ve 250 epok sürmüştür. Daha uzun eğitimin karşılaştırma sonucunu etkilemediği görülmüştür. Doğrulama kümesinde en iyi sınıflandırma başarısı epoğunda, test kümesi başarıları kaydedilmiş, ortalamaları, standart sapmaları ve maksimum değerleri kaydedilmiştir. Sonuçlar Tablo 2 de verilmiştir.

İlk olarak basit üç katmanlı ağlar karşılaştırılabilir. Tablo 2 ilk iki sırada basit odaklı katman ağı (B.OA), basit tam bağlı katman ağı (B.TBA) ait sonuçlar incelenirse odaklı ağın tam bağlı katmana göre dört veri kümesinde de daha iyi ortalama ve en yüksek sınıflandırma doğruluğu elde ettiği görülebilir. MNIST ve FASHION kümelerinde performans farkları küçüktür. Bunun yanında MNIST-cluttered ve CIFAR10 gibi görece daha zor veri kümelerinde fark daha yüksektir. Yaptığımız iki örnekle t-test ile MNIST-CLT ve CIFAR10 daki farkların istatistiksel olarak anlamlı ($p<7e-11$ ve $p<2e-7$) ve odaklı ağ lehine olduğu gözlenmiştir.

Tablo 2' de son iki sırada ise evrişimsel ağlar karşılaştırılmıştır. Evrişimsel ağların performansı beklenildiği gibi basit ağlara göre daha yüksektir. Kendi aralarında karşılaştırıldığında MNIST ve FASHION'da tam bağlı evrişimsel ağ (EV.TBA), MNIST-CLT ve CIFAR-10 veri kümelerinde ise odaklı evrişimsel ağ lehine sonuçlar gözlenmektedir.



Şekil 8. MNIST veri kümesinde tam bağlı ağ (TBA) ve odaklı ağ (OA) karşılaştırması. a) Saklı katmanda yer alan 40 odağın eğitim sonundaki durumları. b) Merkezlerin eğitim sırasında değişimi. c) Eğitim sırasında eğitim ve doğrulama kümesinde hata. d) doğrulama kümesinde sınıflandırma yüzdesi.

(Comparison of a simple 2-layer focusing network with fully connected network in MNIST dataset. a) Change of the focuses in the hidden layer, b) change of the centers, c) and d) loss and classification accuracy comparison.)

Tablo 2. Genel imge veri kümelerinde sınıflandırma başarıları. B.OA basit odaklı ağ, B.TBA basit tam bağlı ağ, EV.OA evrişimsel odaklı ağ, EV.TBA evrişimsel tam bağlı ağ. Her hücre bağımsız sınama kümesinde beş tekrarda elde edilen doğru sınıflandırma yüzdesi ortalaması, standard sapması ve en yüksek değerini göstermektedir.

(The classification results in common image data sets. B.OA simple network with focusing layer, B.TBA simple fully connected network, EV.OA convolutional network with focusing layer, EV.TBA convolutional network with fully connected layer. Each cell shows the mean and standard deviation (and maximum) of classification results in an independent test set.)

| Ağ | MNIST | MNIST-CLT | CIFAR10 | FASHION |
|---------------|--------------------------|--------------------------|---------------------------|--------------------------|
| B.OA | 99.13±.06 (99.21) | 72.24±.33 (72.74) | 61.76±.3 (62.26) | 90.8±.2 (91.2) |
| B.TBA | 98.96±.04 (99.03) | 63.59±.21 (63.98) | 58.99±.18 (59.29) | 90.57±.12 (90.69) |
| EV.OA | 99.52±.05 (99.61) | 94.05±.29 (94.54) | 73.96±.002 (74.16) | 93.01±.14 (93.23) |
| EV.TBA | 99.54±.05 (99.61) | 93.19±.14 (93.39) | 72.49±.002 (72.9) | 93.21±.13 (93.37) |

Yaptığımız iki örnekli t- test ile MNIST-CLT ve CIFAR10 daki farkların istatistiksel olarak anlamlı ($p < 0.003$ ve $p < 0.023$) olduğu gözlenmiştir.

4. TARTIŞMA VE SONUÇ (DISCUSSION AND CONCLUSION)

Bu makalenin araştırma sorularından ilki uyarlanırlar yerel bağlı (odaklanan) nöronların anlamsız girdilerden uzaklaşarak anlamlı girdilere yönelip yönelemeyeceği idi.

Sentetik veri kümelerinde, anlamlı veriye yönelim gözlenmiştir. Hata işlevinde bir değişiklik olmamasına rağmen nöronların anlamlı veriyi kapsayacak şekilde konumlandığı ve başlangıçta görece dar olan odakların eğitim ile daha çok veriyi kayıpsız toplayabilmek için açıldığı gözlenmiştir.

İkinci soru ise uyarlanırlar yerel bağlantı yeteneğinin bir ağa sınıflandırmada avantaj sağlayıp sağlayamayacağı idi. Farklı koşullarda rastgele sentezlenmiş 750 farklı veri

kümesi üzerinde eğitip sinayarak, odaklı katmanların anlamsız veri eklenmiş kümelerdeki sınıflandırma başarısının tam bağlı ağa göre çoğunlukla daha iyi olduğu gözlenmiştir. Yine aynı deneyle, odaklı nöronlardan oluşan bir katmana sahip ağ eğitmenin tam bağlı ağa göre ortalama 1.3 kat daha uzun sürdüğü tespit edilmiştir.

Popüler imge tanıma veri kümelerinden MNIST, MNIST-CLT, CIFAR10, FASHION veri kümeleri üzerinde yapılan deneylerde, odaklı katman kullanmanın özellikle zorlu veri kümelerinde sınıflandırma performansını olumlu etkileyebileceği gösterilmiştir.

Bu makalede yer alan deneylerde odaklanan nöronun sınıflandırma başarısı tam bağlı nöron ile karşılaştırılmıştır. Ancak, odaklı katmanlı ağların da çıkış katmanlarında tam bağlı nöronlar kullanılmıştır. Çünkü, odaklı nöron girdi uzayını yerel parçalara ayırırken, tam bağlı bir katmanın yerel parçalardan gelen temsili bilgiyi tümüyle kullanması gereklidir. Dolayısıyla biyolojik yapılarda olduğu gibi karmaşık ve çoklu problemleri çözmek için farklı nöron tipleri gerekli olabileceğinden birbirlerine üstünlükleri kullanım yerine göre değişebilir.

Bu çalışmada incelenen model bir boyutlu topoloji için geliştirilmiştir. İki veya üç boyutlu türevlerin imge tanıma gibi problemlerde daha iyi sonuçlar elde etmesi beklenir. Bunun yanında, odak parametrelerin eğitim sırasında sabitlenmeyerek, çalışma sırasında dinamik olarak kontrol edildiği bir model ile güçlendirilmiş öğrenmede ve/veya özyineleyen ağlarda kullanılması mümkün olabilir. Çalışmalarımız çok boyutlu odaklanan nöronların geliştirilmesi ve derin ağ uygulamaları ile genişletilecektir.

TEŞEKKÜRLER (ACKNOWLEDGEMENTS)

Bu çalışma BAP 16A202 kapsamında 2016 Haziran 2017 Haziran arasında Işık Üniversitesi tarafından, 118E722 no'lu proje kapsamında Şubat 2019'dan itibaren TÜBİTAK tarafından desteklenmiştir. Ayrıca bazı deneyler NVIDIA-Akademik bağış kapsamında verilen Tesla K40 gpu üzerinde gerçekleştirilmiştir. Modelin ilk çalışmalarında ve Python dili ile kodlanmasında İlker Çam'nın önemli katkıları olmuştur. Modelin geliştirilmesi sırasında Dr. Deniz Karlı, Dr. Mehmet Önal, Dr. Murat Ayhan, Dr. Olcay T. Yıldız, Dr. Emine Ekin değerli fikirlerini bizimle paylaşmıştır.

KAYNAKLAR (REFERENCES)

- [1] Y. Sun et al., "Deep-id3: Face recognition with very deep neural networks", *CoRR, arXiv:1502.00873*, 2015.
- [2] Y. Zhang et al., "Very deep convolutional networks for end-to-end speech recognition", *ICASSP*, Lousianna, 4845-4849, 2017.
- [3] Tai et al., "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks", *CoRR, arXiv:1503.00075*, 2015.

- [4] D. Silver, A. Huang, et al., "Mastering the game of Go with deep neural networks and tree search", *Nature*, 529, 484-489, 2016.
- [5] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, 521, 436-444, 2015.
- [6] C. Cortes, X. Gonzalvo, V. Kuznetsov, M. Mohri, S. Yang, "Adanet: Adaptive structural learning of artificial neural networks", *ICMLR*, 70, 874-883, Sydney, 2017.
- [7] E. Fiesler, "Comparative bibliography of ontogenic neural networks", *ICANN*, 793-796, Springer, 1994.
- [8] B. Hassibi, D. G. Stork, G. J. Wol, "Optimal brain surgeon and general network pruning", *IEEE Int. Conf. on Neural Networks*, 1, 293-299, 1993.
- [9] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, Y. Bengio, "Fitnets: Hints for thin deep nets", *ICLR*, CA, ABD, 2015.
- [10] B. Baker, O. Gupta, N. Naik, R. Raskar, "Designing neural network architectures using reinforcement learning", *ICLR*, Toulouse, FR, 2017.
- [11] S. Han, J. Pool, J. Tran, W. J. Dally, "Learning both weights and connections for efficient neural networks", *Neural Information Processing Systems*, 1135-1143, Montreal, CAN, 2015.
- [12] A. Coates, A. Y. Ng, "Selecting receptive fields in deep networks", *Neural Information Processing Systems*, Granada, SPN, 2011.
- [13] I. Çam, F. B. Tek, "Odaklanan nöron (focusing neuron)", *IEEE 25th Signal Processing and Communications Applications (SIU)*, 1-4, Zonguldak, TR, 2017.
- [14] E. R. Kandel, *In search of memory: The emergence of a New Science of Mind*, W. W. Norton & Company, 2006.
- [15] A. R. Luria, "The Functional Organization of the Brain", *Scientific American*, 222(3), 66-79, 1970.
- [16] D. J. Graham, "Routing in the brain", *Frontiers in Computational Neuroscience*, 8, 44, 2014.
- [17] C. D. Gilbert, W. Li, V. Piech, "Perceptual learning and adult cortical plasticity", *The Journal of Physiology*, 30, 2743-2751, 2009.
- [18] T. Suter, Z. J. DeLoughery, A. Jaworski, "Meninges-derived cues control axon guidance", *Developmental Biology*, 430, 1-10, 2017.
- [19] S. J. Pan, Q. Yang, "A survey on transfer learning", *IEEE Trans. on knowledge and data engineering*, 22(10), 1345-1359, 2010.
- [20] A. Soltoggio, K. O. Stanley, S. Risi, "Born to learn: the inspiration, progress, and future of evolved plastic artificial neural networks", *Neural Networks*, 108, 48-67, 2018.
- [21] B. Baker, O. Gupta, N. Naik, R. Raskar, "Designing neural network architectures using reinforcement learning", *ICLR*, Toulon, FR, 2017.
- [22] H. Liu, K. Simonyan, Y. Yang, "Darts: Differentiable architecture search", *CoRR, arXiv 1806.09055*, 2018.
- [23] M. T. Hagan, H. B. Demuth, M. H. Beale, *Neural Network Design*, Martin Hagan, 2014.
- [24] D. Elizondo, R. Fiesler, "A survey of partially connected neural networks", *Int J. Neural Systems*, 8, 535-568, 1997.

- [25] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, "Deepface: Closing the gap to human-level performance in face verification", *CVPR*, 1701-1708, 2014.
- [26] H. A. Rowley, S. Baluja, T. Kanade, "Neural network-based face detection", *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 23-38, 1998.
- [27] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", *Proc. of the IEEE*, 86, 2278-2324, 1998.
- [28] K. Gregor, Y. LeCun, "Emergence of complex-like cells in a temporal product network with local receptive fields", *CoRR*, *arXiv:abs/1006.0448*, 2010.
- [29] T. Poggio, T. Serre, "Models of visual cortex", *Scholarpedia*, 8, 4, 3516, 2013.
- [30] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain, cornell aeronautical laboratory", *Psychological Review*, 65, 386-408, 1958
- [31] B. A. Olshausen, D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images", *Nature*, 381, 607-609, 1996.
- [32] S. Munder, D. M. Gavrila, "An experimental study on pedestrian classification", *IEEE Trans. Pattern Anal. Mach. Int.*, 28, 1863-1868, 2006.
- [33] M. J. L. Orr, Introduction to radial basis function networks, **Report Gatech**, 1996.
- [34] T. Kohonen, "The self-organizing map", *Proceedings of the IEEE*, 78, 1464-1480, 1990.
- [35] E. Oja, "Simplified neuron model as a principal component analyser", *Journal of Mathematical Biology*, 15, 267-273, 1982.
- [36] T. Miconi, J. Clune, K. O. Stanley, "Differentiable plasticity: training plastic networks with gradient descent", **ICML**, Stockholm, Sweden, 2018.
- [37] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans. Pattern Anal. Mach. Int.*, 20, pp 1254 1259, 1998.
- [38] B. Olshausen, C. Anderson, D. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information", *Journal of Neuroscience*, 13, 4700-4719, 1993.
- [39] K. Xu, J. L. Ba, R. K. et al., "Show, attend and tell: Neural image caption generation with visual attention", **ICML**, 37, 2048-2057, 2015.
- [40] J. Ba, V. Mnih, K. Kavukcuoglu, "Multiple object recognition with visual attention", *CoRR*, *arXiv 1412.7755*, 2014.
- [41] B. Cheung, E. Weiss, B. A. Olshausen, "Emergence of foveal image sampling from learning to attend in visual scenes", *CoRR*, *arXiv: abs/1611.09430*, 2016.
- [42] S. Sabour, N. Frosst, G. E. Hinton, "Dynamic routing between capsules", *CoRR*, *arXiv: abs/1710.09829*, 2017.
- [43] Theano Development Team, "Theano: A {Python} framework for fast computation of mathematical expressions", *CoRR*, *arXiv:abs/1605.02688*, 2016
- [44] Internet: <https://github.com/btekgit/FocusingNeuron>.
- [45] F. B. Tek, "An Adaptive Locally Connected Neuron Model: Focusing Neuron", *CoRR*, *arXiv:1809.09533*, Aug, 2018.
- [46] B. Can, "LSTM Ağları ile Türkçe Kök Bulma". *Bilişim Teknolojileri Dergisi*, 12(3), 183-193, 2019
- [47] M. A. Kızrak, B. Bolat, "Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma". *Bilişim Teknolojileri Dergisi*, 11(3), 263-286, 2018.