



Anlamsal Boşluk Doldurulmasında Derin Öğrenme Yöntemlerinin Kullanılması Üzerine Bir İnceleme

İbrahim Ali Metin^{1*}, Bahadır Karasulu²

¹Çanakkale Onsekiz Mart Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Çanakkale, Türkiye

²Çanakkale Onsekiz Mart Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Çanakkale, Türkiye

Özet

Anlamsal boşluk kavramı, makineler aracılığıyla veriden elde edilen temel renk, doku ve şekil gibi özniteliklerle insan tarafından aynı veriye bakıldığında algı yoluyla tanımlanan kavramsal sonuçların farklılıklarından doğmaktadır. Bu boşluğun giderilmesi için çeşitli yol ve yöntemler literatürdeki çalışmalarda denenmiştir. Bu çalışmalarda, ontolojik sistemlerin geliştirildiği ve arama işlemlerinin de bu tarafa yoğunlaştırılmasının sağlandığı görülmektedir. Çalışmamızın ana amacı genel olarak Anlamsal Tabanlı Görüntüden Bilgi Elde Etme çalışmalarıyla İçerik Tabanlı Görüntüden Bilgi Elde Etme için gerçekleştirilen aramada ortaya çıkan Anlamsal Boşluk sorunun üstesinden gelinmesine dair yapılan çalışmaları ortaya koymaktır. Çalışmamızda literatürdeki çeşitli uluslararası dergi ve konferans yayınları incelenerek, kıyaslamalı özlü ve sistematik bir literatür taraması sunulmaktadır. Bilimsel tartışmaya da yer verilmektedir.

Makale Bilgisi

Başvuru:
30/04/2024
Kabul:
26/06/2024

Anahtar Kelimeler: Yapay Zekâ, Görüntü İşleme, Anlamsal Boşluk, Derin Öğrenme, Evrişimli Sinir Ağı

A Review On Using Deep Learning Techniques To Bridge The Semantic Gap

Abstract

The concept of semantic gap arises from the differences between the basic attributes such as color, texture and shape extracted from the data by machines and the conceptual results defined by human perception when looking at the same data. Various ways and methods have been tried in the literature to bridge this gap. In these studies, it is seen that ontological systems have been developed and search operations have been focused in this direction. The main purpose of our study is to present the studies on Semantic Based Image Retrieval in general and the studies on overcoming the Semantic Gap problem that arises in the search for Content Based Image Retrieval. We present a concise and systematic comparative literature review by analyzing various international journal and conference publications in the literature. Scientific discussion is also included.

Keywords: Artificial Intelligence, Image Processing, Semantic Gap, Deep Learning, Convolutional Neural Network

* İletişim e-posta: ibrahimalimetin@gmail.com

1 Giriş

Bilgisayar sistemlerinin gelişen teknoloji ile günden güne ilerlemeleri ve bu ilerleyişin sonucunda ortaya çıkarılan daha başarımlı yüksek sistemler sayesinde mevcut veri kümelerine kıyasla daha çok veri barındıran veri kümeleri kullanılarak işlem yapılabilir duruma gelmiştir. Bahse konu bu büyük veri kümelerinin kullanılması, belirli ön işlemler uygulanmış bu veriler üzerinden anlamlı kullanıma uygun bilgilerin elde edilmesi ve bu anlamlı bilgiler ile gelişmiş karar sistemlerinin oluşturulması makine öğrenimi teknikleri ile daha etkin ve kolay olmaktadır. Makine öğrenimi yöntemleri kullanılan cihazlar ya da bilgisayar tarafından tecrübe edinilen problemlerin çözümüne ilişkin yaklaşımlar sayesinde, çözümün daha iyiye gitmesi veya gelişim için ne şekilde bir yol izleneceği konuları da ele alınmaktadır. Bu durum, genel problem çözümünde temel amaçlar olarak da ifade edilebilir [1]. Görüntü verilerinden oluşan bir veri kümesi içerisinde yer alan nesnelere algılayabilmek için günümüz teknolojik gelişmelerinde Anlamsal Tabanlı Görüntüden Bilgi Elde Etme (Sematic-based Image Retrieval, SBIR) alanında çalışmalar devam etmektedir. Bahsi geçen bu çalışmalar yardımıyla anlamsal boşluk kavramı ile makineler kullanılarak görüntü veri kümeleri üzerinden elde edilen; renk, doku ve biçim gibi öznitelikler ile insan tarafından görüntü veri kümeleri üstünden algı sayesinde tanımlanabilen kavramların birbirleri ile uyumsuzluklarını ortadan kaldırmak üzerine yoğunlaşmıştır. Bu amaç doğrultusunda yola çıkılarak oluşturulan ontolojik sistemler oluşturulmuş ve arama işlemlerinin de bu alanda yoğunlaştırılması sağlanmıştır. Görüntüden bilgi elde etme, kullanıcının sorgulama amacıyla kullanmakta olduğu görüntüye erişilerek benzer görüntüler için veri tabanında arama yapmayı içerir. Görüntülere genellikle üç farklı yöntem sayesinde erişilebilmektedir. Bu yöntemlerden ilki metin tabanlı görüntüden bilgi elde edilmesi, ikinci yöntem ise içerik tabanlı görüntüden bilgi elde edilmesi ve son olarak üçüncü yöntemde anlamsal tabanlı görüntüden bilgi elde edilmesi denir [2]. Arama işlemlerinde görselin içeriğini belirten anahtar kelimelere göre arama işlemi yapıldığında, aranan kelime ile ilgili olmayan görseller de sonuç olarak gösterilebilmektedir. Metin tabanlı yoldan bir görüntüye erişmeden önce, görüntünün içerisindeki içerik hakkında anlamlı açıklama sağlayan metinsel bir açıklamayı el ile veya otomatik şekilde eklemek gerekmektedir. El ile

oluşturulan açıklama, yoğun çaba ve zaman gerektiren bir işlem adıdır. Hatalara açıktır. Alternatif olarak, otomatik altyazı tekniği kullanılarak içeriğe en uygun olacak açıklamaların eklenmesi için literatürde çeşitli araştırmalarda çözüm yöntemleri önerilmiştir [3]. Görüntü verisini alma sürecini iyileştirmeye yönelik Ngo ve diğerleri tarafından 2016 yılında gerçekleştirilen literatürdeki çalışma Destek Vektör Makinesi (Support Vector Machine, SVM) öğrenimi yöntemi ve alakalı geri bildiriminin bir kombinasyonu kullanılmıştır. Arama için kullanılanlara benzer olarak görüntüler, SVM kullanılarak veri tabanından elde edilmiştir. İlgili durum geri bildirim yoluyla kullanıcı ile sistem etkileşimi sayesinde, algoritmalar, anlamsal boşluk kavramsal tanımını olabildiğince aza indirmek için veri tabanından elde edilen sonuçlara dayalı olarak kullanıcıların verdikleri karar doğrultusunda eğitilmiştir [4]. Bölütleme kavramı, görüntüdeki her bir piksel bilgisine bir sınıfın etiketini atamaya dayalı bir modeli ifade etmektedir. Anlamsal olarak ifade edilen görüntülerle ilişkili piksel tabanlı bilgiler, nesne tanıma ve karar verme işlemlerinde kullanılmaktadır. Evrişimli Sinir Ağı (Convolutional Neural Network, CNN) mimarisi, daha etkili başarımlı sonuçlarını sağlayan bölütleme için kullanılan bir model olarak literatürde kendine yer bulmuştur [5].

Derin öğrenme mimari modellerinin geliştirilmesi doğrultusunda makine öğrenimi konusunda gerçekleştirilen araştırmalar büyük bir aşama kaydetmiştir. Derin öğrenme mimarilerinin amacı, bilgisayarların önerilen yazılımsal mimari modelleri kullanarak algoritmik bir yoldan daha yüksek düzeyde temsil edilebilen ve insanların görüntülerden veya ilgili verilerden algıladığı kavramları otomatik olarak öğrenmesidir. Öğrenme aktarımı (Transfer Learning), bir makine öğrenimi tekniği, tıpkı insanların yaptığı gibi, bir problemi çözerken elde ettiği bilgileri saklamakta ve bu bilgiyi başka bir problemle karşılaştığında kullanılmaktadır. İlgili yazılımsal modeli elde etmek için önceki bilgileri kullanan öğrenme aktarımı, böylece daha az eğitim verisiyle daha büyük başarı gösterebilir ve daha hızlı öğrenebilir [6]. Temsili öğrenme (Representation Learning) sayesinde düşük seviyeli öznitelikler (renk, doku, biçim gibi) kullanılarak daha yüksek seviyeden öznitelikler (anlamsallık içeren) elde edilebilmekte ve öğrenme kalitesi artırılmaktadır.

Çalışmamızda literatürde yer alan farklı uluslararası dergi makaleleri ve konferans bildirileri incelenerek, aşağıdaki bölümlerde bunların detayları verilen biçimde karşılaştırmalı olarak sunulmaktadır. Çalışmamız dört bölüme ayrılmıştır. İkinci bölüm olan "Önceki Çalışmalar" bölümünde literatürdeki uygulamalı araştırmalara temel oluşturan yöntemler ve çeşitli teknik altyapılardan bahsedilmektedir. Üçüncü bölümde çalışmamızdaki araştırma sonucunda elde edilen bilimsel bulgulara yer verilirken, dördüncü bölümde ise çalışmamızdaki bilimsel bulgulara dayanan tartışma ve sonuçlara yer verilmektedir.

2 Önceki Çalışmalar

Bu bölümde literatürde yer alan önceki çalışmalar incelenmektedir. Temel yöntemimiz bilgi kaynaklarının sistematik olarak araştırılarak uygun kategoriler altında benzer çalışmaların belirli özelliklerine göre gruplandırılmasına dayanmaktadır. Literatür incelendiğinde, Tzelepi ve Tefas'ın 2018 yılında gerçekleştirdiği bir çalışmada, derin CNN mimarisini tekrardan eğitmeye dayalı, veri kümesine ilişkin hiç bir veri içermeyen, tamamen denetimsiz bir yeniden eğitim yaklaşımıyla verilerin sınıflarını belirleyen etiketleri varsa, mevcut sınıf bilgisine ait etiketleri kullanılmaktadır. Kullanıcı etkileşimi gerçekleştirildiğinde, ilgili geri bildirim ve sonuçlara dayalı kullanıcı kararları kullanılarak algoritmanın tekrardan eğitilmesi gerçekleştirilmiştir [7].

CNN kullanan başka bir çalışmada öznetellik elde etme, iki değişik CNN mimarisinin kullanılması ile paralel olarak gerçekleştirilmiştir ve Oxford 5K [8] veri kümesinde %95 olan yüksek başarı oranı elde edilmiştir [9]. Noh ve arkadaşları tarafından 2015 'de yapılan bir çalışmada; sunulan evrişimli-dekonvolüsyonel (ters evrişim işlemi) tabanlı mimariyi kullanan bir modelde, Tam bağlantılı Evrişimli Ağ'da karşılaşılan dezavantajları ele almak için, bu tip ağlar kullanılarak öznetellik elde etme ve ters evrişimli ağlar kullanılarak elde edilen bu öznetelikleri kullanarak nesne bölütleme için bir örnek teşkil edecek çalışma gerçekleştirilmiştir [10]. Badrinarayanan tarafından 2017 'de sunulan başka bir çalışma ele alınmış bölütleme tabanlı ağ olan SegNet mimarisi, piksel tabanlı anlamsal bölütleme gerçekleştirmekte ve bir kodlayıcı (encoder), bir kod çözücü (decoder) ve bir piksel tabanlı sınıflara ayrıştırma katmanından bir araya gelmektedir. Alta yatan Görsel Geometri Grup

(VGG-16) isimli 16 katmanlı derin sinir ağı mimarisinin tamamen bağlantılı katmanı kaldırılarak ve bu katmanın alternatifi olarak piksel tabanlı bir sınıflara ayrıştırma katmanı eklenerek ölçek olarak daha küçük ve karmaşıklık göz önüne alındığında hız bakımında iyi durumda bir mimari öne çıkarılmıştır [11].

Panoptic bölütleme yardımıyla görüntü verisine ait her piksel bir sınıf etiketi ile işaretlenmektedir. Gerçekleştirilmiş bir çalışma öne sürülen Panoptik Bölütleme Ağı (Fast Panoptic Segmentation Network - FPSNet) kullanılarak görüntü verisinin içerdiği öznetelikler varlık ve soyut sınıflar benzeri farklı sınıflara ayrılarak bölütleme işlemi gerçekleştirilmektedir [12]. Bir diğer çalışmada; Çizge tabanlı bölütleme, yüksek değişkenlikli bölgelerdeki ayrıntıları görmezden gelirken düşük değişkenlikli görüntü bölgelerindeki ayrıntıları koruma yeteneğine sahip olduğu görülmüştür. Bölütlemeyen sonra, segmentli görüntüler için özneteliklerin elde edilmesini sağlar, doku öznetelikleri için dalgacık dönüşümü kullanılarak aynı zamanda renk öznetelikleri içinde histogram modeli kullanılarak sonrasında segmentli sorgu görüntü öznetelikleri segmentli veri tabanı görüntülerinin öznetelikleriyle karşılaştırılır. Doku öznetelikleri için kullanılan benzerlik ölçüsü Öklid uzaklık ölçüsüdür ve renk öznetelikleri için Karesel uzaklık yaklaşımıdır. Deneysel sonuçlar incelendiğinde, bölütlemeyle renk özneteliği için performansta yaklaşık %12'lik bir iyileştirme olduğunu göstermektedir [13]. Özden ve Ediz tarafından 2006 yılında gerçekleştirilen bir başka çalışmada ise renk, doku ve mekansal bilgi gibi düşük seviyeli özneteliklere dayalı görüntü segmentasyonu için yeni bir yaklaşım sunulmuştur. Önerilen yaklaşım, ortalama kaydırma algoritmasında filtreleme için öznetelik alanını genişletmeye dayanmaktadır. Bu yöntem, farklı yönelimlerdeki doku enerjisine dayalı olarak dokulu bölgeleri önceden tanımlanmış doku sınıflarına karakterize etmek için ayrık dalgacık çerçeveleri kullanır. Bu yaklaşımın performansı, renk ve doku bilgisinin mevcut olduğu doğal görüntüler kullanılarak gösterilmiştir. Sonuçlar, önerilen yaklaşımın standart ortalama kaydırma filtrelemesinden daha sağlam ve doğru olduğunu göstermektedir [14]. Araştırılan "Bölütleme Tabanlı Görüntüden Bilgi Elde Etme" konusunda birçok çalışma kaynaklarında ele alınan teknikler, veri kümeleri, algoritmalar, başarımlar değerlendirme

ölçütleri ve bu başarımların sonuçları Tablo 1'de gösterilmiştir.

Tablo 1. "Bölütleme Tabanlı Görüntüden Bilgi Elde Etme" Konusundaki Farklı Araştırma Çalışmaları ve Çalışmalardan Edinilen Başarımların Sonuçları

Makale	Yöntem	Veri Kümesi	Başarım Ölçütü	Başarım Sonuçları
Long ve diğerleri, 2015 [5]	Evrişimli Sinir Ağları	PASCAL VOC NYUDv2 SIFT Flow	Ortalama Birleşim (Mean IU)	%62,6 %34,0 %39,5
Noh ve diğerleri, 2015 [10]	Evrişimli Sinir Ağları	PASCAL VOC 2012	Birleşim üzerinde kesişim (Intersection over Union IoU)	%72,5
De Geus ve diğerleri, 2020 [12]	Hızlı Panoptik Bölütleme Ağı	PASCAL VOC 2012 Cityscapes	Panoptik kalite ölçüğü (Panoptic Quality Metric)	%54,8 %55,1
Suhasini ve diğerleri, 2008 [13]	Çizge Tabanlı Bölütleme	Görüntü Veri Kümesi	Öklid mesafe ölçümü (Euclidean distance measure)	%12
Özden ve Ediz, 2006 [14]	Ayrık Dalgacık Çerçevesi	Görüntü Veri Kümesi	Standart ortalama kaydırma filtresi (standard mean-shift filtering)	Başarılı

Derin öğrenme mimari modelleri, özellikle de CNN mimarisi, görüntü bölütleme ve sınıflandırma işlemleri için yaygın olarak literatürdeki çalışmalarda yer almaktadır. CNN ve tam bağlantılı evrişimli ağ mimarileri, tıbbi görüntülerde derin öğrenme yaklaşımları kullanılarak görüntü bölütleme için yüksek başarımlara sahip sonuçları sağlamaktadır. Haque ve Neubert tarafından 2020 yılında gerçekleştirilen bir çalışmada; tıbbi görüntüleme işleminden sonra yapılan görüntü bölütleme için ele alınan derin öğrenme tekniklerini özetlemektedir [15]. Literatürdeki çeşitli çalışmalarda, görüntü verisinin içerdiği içerik ile yine bu görüntü verisini tanım olarak ifade eden sınıf etiketleri arasındaki anlamsal boşluk sorununu çözmek için geliştirilen çerçevedeki görüntüden yapılan çıkarsamada öznitelikleri kullanarak bir görüntü benzerlik diyagramı ortaya çıkarılmıştır. Bu sayede görsel içeriği ile ana görseli tanım etiketini ortaya çıkartmak için ele alınacak

sınıflandırma etiketlerinin aralarındaki anlamsal boşluk olarak ifade edilen uçurumun kapatılmasında etkili sonuçlar elde edildiği gözlemlenmiştir [16]. Anlamsal boşluk problemi, sığ ağ'ın katman öznitelikleri ile birlikte derin ağ'ın katman öznitelikleri arasında ortaya çıkar. Bu belirtilen sorunu gidermek adına ortaya atılan yöntemler, paralel piramitler ve anlamsal bölütleme için farklı düzeylerde öznitelik kullanmak için aşağıdan yukarıya (bottom-to-top) bir yaklaşım şekli denenmiştir [17]. Araştırılan "İçerik Tabanlı Görüntüden Bilgi Elde Etme" konusunda birçok çalışma kaynaklarında ele alınan teknikler, veri kümeleri, algoritmalar, başarımlar değerlendirme ölçütleri ve bu başarımların sonuçları Tablo 2'de gösterilmiştir. Bu çalışmalarda sırasıyla; Corel veri kümesi, Paris 6K [18] veri kümesi, UKBench [19] veri kümesi ve Oxford 5K [8] veri kümesi kullanılmıştır.

Tablo 2. "İçerik Tabanlı Görüntüden Bilgi Elde Etme" Konusundaki Farklı Araştırma Çalışmaları ve Çalışmalardan Edinilen Başarım Sonuçları

Makale	Yöntem	Veri Kümesi	Algoritma	Başarım Ölçütü	Başarım Sonuçları
Ashraf ve diğerleri, 2018 [28]	Yapay Sinir Ağları	Corel	Ayrık Dalgacık Dönüşümü	Ortalama Duyarlık (Mean Precision)	%73,5
Tzelepi ve Tefas, 2018 [7]	Evrışimli Sinir Ağları	Paris 6K, UKBench	Uygulanabilir Moment Tahmin Algoritması	Ortalama Doğruluk (Mean Accuracy)	Denetimli (%98,59) Denetimsiz (%83,47)
Alzu'bi ve diğerleri, 2017 [9]	Evrışimli Sinir Ağları	Oxford 5K	İki Çizgili Model	Ortalama Doğruluk (Mean Accuracy)	%95,7
Yenigalla ve diğerleri, 2023 [29]	Derin Öğrenme	Görüntü Veri Kümesi	ResNet50 VGG16	Duyarlık (Precision)	%100
Sikandar ve diğerleri, 2023 [30]	Makine Öğrenmesi	Corel 1K, 5K, 10K	Destek Vektör Makinesi K En Yakın Komşuluk Evrışimli Sinir Ağları	Duyarlık (Precision)	%80 %60 %50

Anlamsal boşluk alanında gerçekleştirilen çalışmalar incelendiğinde farklı yöntemlerin farklı veri kümeleri üzerindeki gösterdikleri başarımlarının olduğu görülmektedir. Zhao ve Grosky'nin 2002 yılındaki çalışmasında gizli anlamsal indeksleme kullanılarak Web üzerinden elde edilen verilerle yapılan deneylerde Duyarlık (Precision) %35 ve Anma (Recall) değeri %88 olarak elde edilmiştir [19]. Ma ve diğerleri'nin 2010'daki çalışmasında Flickr veri kümesi [16]. üzerinde Rastgele Dizilim Modeli ile yapılan deneylerle N-Duyarlık (Precision@N) değeri olarak %68 elde edilmiştir. Flickr oldukça büyük bir veri kümesi oluşturulmasında etkin kullanılmıştır [20, 21]. Pang ve diğerlerinin 2019 yılındaki çalışmasında ise, Paralel Özellik Piramidi yöntemi kullanılarak Ortalama Birleşim Üzerinde Kesişim (Mean IoU) ölçüt değeri olarak sırasıyla; Pascal VOC 2012 [17, 22] veri kümesi için %88,6, CamVID (Brostow ve diğerleri, 2008) veri kümesi için %68,4 ve Cityscapes [23] veri kümesi için ise %81,2 değeri elde edilmiştir. Scopus [24], Web of Science [25] ve IEEE [26] veri tabanlarındaki çalışmalara bakıldığında; öznitelik elde etme, Anlamsal Tabanlı Görüntüden Bilgi Elde Etme yönteminde önemli bir role sahiptir. Öznitelik elde etmede en çok tercih

edilen mimari model Evrışimli Sinir Ağlarıdır. Ayrıca ilgilenilen nesnelere tanımlanmasında farklı bilgi alanlarının örneğin WordNet [27] kullanımı da yaygındır. Ashraf ve diğerlerinin 2018 yılındaki çalışmasında yapay sinir ağı ve ayrık dalgacık dönüşümü kullanılmıştır [28]. Gerçekleştirilen bir diğer çalışmada transfer öğrenme tekniğini kullanan ve iki önceden eğitilmiş derin öğrenme modeli ResNet50 ve VGG16 ile bir makine öğrenme modeli kullanılarak uygulanan yeni bir hibrit derin öğrenme ve makine öğrenmesi tabanlı CBIR sistemi sunulmuştur. Benzer görüntülerin sonucunu göstermek için bir web arayüzü oluşturulmuş ve %100'e ulaşan başarımlar gözlemlenmiştir [29]. Sikandar ve diğerlerinin 2023 yılında gerçekleştirdiği farklı bir çalışmada ise Destek Vektör Makinesi, K En Yakın Komşuluk ve Evrışimli Sinir Ağları olmak üzere üç makine öğrenme yöntemini kullanan CBIR sistemini, verileri %80 eğitim verisi ve %20 test verisi olarak bölerek Corel 1K, 5K ve 10K veritabanlarını kullanarak işlemişlerdir. Sonuç olarak, kullanılan algoritmalarının görüntüden bilgi elde etme görevini tamamlamada ne kadar etkili olduğuna dair net bir vizyon sağlamaktır [30]. Song ve diğerlerinin 2018 yılında gerçekleştirdikleri

çalışmada Evrişimli Sinir Ağları kullanılarak görüntü veri kümelerinin öznelik elde etmeyi gerçekleştirilerek geliştirilen kayıp fonksiyonları sayesinde bu ağlar eğitilerek parametreler optimize edilmiş ve sınıflandırma doğruluğunda iyileşmelerin olduğu gözlemlenmiştir [31]. Farklı derin mimari modellerinin birlikte kullanılması sayesinde görüntülerden algılanan yüksek düzeydeki anlamsal özneliklerin elde edilmesi alanında geliştirmeler yapılmaktadır. Yüksek düzeydeki özneliklerin elde edilmesi amaçlı Wu tarafında yapılan 2020 yılındaki bir çalışmada; Derin Boltzman Makineleri ile Evrişimli Sinir Ağları beraber kullanılmıştır. Evrişimli Sinir Ağları ile düşük düzeydeki özneliklerin elde edilmesi sağlanmış, Derin Boltzman Makineleri ile görüntüye ait yüksek düzeydeki ilişkiler oluşturulmuştur [32]. Literatürdeki bir başka çalışmada Bouchakwa ve diğerleri Çok Seviyeli Görüntü Çeşitlendirme Süreci şeklinde adlandırılan teknik sayesinde, anlamsal olarak benzer resimlerin aynı veri kümesi içerisinde toplanmış ve sorgu ifadesi göz önüne alınarak filtrelenip yeniden sıralanması gerçekleştirilmiştir [33]. Farklı bir araştırmada öne sürülen "OntoLook" isimli sistem sayesinde web sayfalarının içeriği, anlamsal ağda kullanılan ifadeler ile (Web Ontology Language - OWL, Resource Description Framework - RDF) yapılandırılmış ve anahtar kelime temelli bir arama motoru ortaya çıkarılmıştır [34]. Yapılan farklı bir çalışmada ise çiçek bilgi alanı üzerinde Çok Seviyeli Görüntü Ontolojisi (Multi-level image ontology) yaklaşımı sayesinde yazı ile görsel özneliklerin kombinasyonu ile düşük seviyedeki özneliklerin anlamlandırılması gerçekleştirilmiştir [35].

3 Bulgular

Bu bölümde, literatür taramasının ardından Scopus [24], Web of Science [25] ve IEEE [26] veri tabanlarından elde edilen çalışmalardan edinilen bulgulardan bahsedilecektir. Literatürde bu çalışmamızda bahsi geçen konuyu ele alırken seçilen araştırma sorusu; "Klasik makine öğrenmesi ile elde edilebilen düşük seviye öznelikler ile daha soyut olan anlamca zengin yüksek seviyeli özneliklerin arasındaki bilgi zenginliğini kısıtlayarak doğruluğu azaltan durumlarda anlamsal boşluğun doldurulmasında derin

öğrenmenin soyutlama, temsili öğrenme ve otomatik öznelik öğrenme yeteneğinin rolü nedir?" olarak belirlenmiştir. Buna istinaden, konu kapsamını ifade edecek anahtar kelimeler olarak; "derin öğrenme (deep learning)", "temsili öğrenme (representation learning)", "anlamsal boşluk (semantic gap)", "görüntü işleme (image processing)" ve benzeri birçok kelime veya sözcük grubu bahsi geçen veri tabanlarının arama motorlarında taranmıştır. Yakın ilişkide olan dergi makaleleri, sözlü bildiriler ve dokümanlar incelenerek içlerinde konuya katkı sunacağı düşünülenler; yayın yıllarına ve konu başlıklarına göre sistematik olarak gruplandırıldıktan sonra aralarından konu bütünlüğü oluşturabilecek olanlar seçilerek bu çalışmamızdaki sistematik incelemenin yapılmasında kullanılmıştır.

Çalışmamızın odak noktası olan "Anlamsal Tabanlı Görüntüden Bilgi Elde Etme" konusunda birçok çalışma kaynaklarında ele alınan teknikler, veri kümeleri, algoritmalar, başarımlar değerlendirme ölçütleri ve bu başarımlar sonuçları Tablo 3'de gösterilmiştir.

Tıbbi alanda içeriğe dayalı görüntüden bilgi elde etme alanında sistemlerin çalışma başarımlarını ve kullanım durumundaki elverişliliğin üzerinde etki yaratan anlamsal boşluklar olarak ifade edilen kavramsallığa ilişkin sistematik bir analiz sistemi ve taksonomi çalışmasının sonucuna ulaşmaya ilişkin kullanılabilirlik hakkında genel bir bakış açısı ortaya atmaktadır. Ayrıca kategorilerin kullanılması da yapılacak olan karşılaştırmalar için kolaylık sağlayacaktır.

Anlamsal Ağ (Semantic Web), son günlerde sıklıkla kullanılmakta olan Web teknolojilerinin ana yaklaşımından, bir belgeler ağından (kullanılan belgelerin birbirleri ile ilişki kurarak kullanılan bu şekilde de aralarında bağlantının sağlandığı bir yaklaşımdır), bir bilgi ağına (belirtilen ya da tanımlanan bir ağ düzlemi üzerinde tanımlanmış ilişki verilerini içeren bilgileri ifade etmektedir) çıkarsamasına dönüşmüştür. Anlamsal ağ yapısı makinelerin işleme alabileceği bir format yapısına bilgiyi dönüştürür ve bu dönüştürdükleri bilgiler arasında anlamsal olarak ilişkilerin kurulmasına olanak sağlamaktadırlar. Bu bakış açısıyla anlamsal ağ oluşturulurken, anlamsal boşluğun olabildiğince kapatılmış olması, tam bir çözüm için gerekli olmaktadır.

Tablo 3. "Anlamsal Tabanlı Görüntüden Bilgi Elde Etme" Konusundaki Farklı Araştırma Çalışmaları ve Çalışmalardan Edinilen Başarım Sonuçları

Makale	Yöntem	Veri Kümesi	Başarım Ölçütü	Başarım Sonuçları
Song ve diğerleri, 2018 [31]	Evrışimli Sinir Ağları	ILSVRC Caltech256	Ortalama Duyarlık (Mean Average Precision - MAP), Anma (Recall)	ILSVRC: %60,8 ve %78,9 Caltech256: %50,2 ve %50,7
Wu, 2020 [32]	Evrışimli Derin Boltzman Makineleri	STL-10	Duyarlık oranı (Precision rate), Anma oranı (Recall rate) F ölçütü (F value)	%63,69, %62,62, %63,15
Bouchakwa, Ayadi, ve Amous, 2020 [33]	Metin Tabanlı Görüntüden Bilgi Elde Etme (İki seviyeli 200 anlamsal terim içerir)	Flickr	Ortalama Duyarlık (Precision) Ortalama Anma (Recall) Ortalama F ölçütü (F value)	%86,86, %83,10, %84,93
Li, Wang, ve Huang, 2007 [34]	Metin Tabanlı Görüntüden Bilgi Elde Etme (Kelime sayısı bazlı)	Web	Zaman karmaşıklığı (Time complexity), Kelime sayısı için CPU süresi bazında Eşik (Threshold) yok, Eşik (Threshold) var	$O(2^n)$, 3125 ms, 500 ms
Minu ve Thyagarajan, 2014 [35]	k -Ortalamlar kümeleme	Hortipedia's Çiçek Veri Kümesi	Doğruluk (Accuracy), Üstün Standart Ölçüm tabanlı (AFIF ontolojisi) Öğrenme Doğruluğu	%72, %92

Anlamsal Tabanlı Görüntüden Bilgi Elde Etme üzerine gerçekleştirilen ve geliştirmeleri devam eden çalışmalar sürdürülmektedir. Bahsi geçen bu çalışmalar yardımıyla anlamsal boşluk ifadesi ile kullanılan makineler yardımıyla görüntü veri kümeleri üzerinden elde edilen renk, doku ve biçim gibi özniteliklerle ele alınan görüntü verileri üzerinden insanın algılama yeteneği yardımıyla tanımlayabilmesi mümkün kavramlar arasındaki uyumsuzluklarının önlenmesi sağlanmıştır.

4 Sonuç

Anlamsallık yapısına dayalı görüntü kümelerinden bilgi elde etmekte asıl hedef, Anlamsal Tabanlı Görüntüden Bilgi Elde Etme altyapısında anlamsal olarak benzerliklerin ortaklaşa gözlemlenebildiği görüntüler mevcuttur. Görüntü özniteliklerinin elde edilmesi ve bölütleme işlemlerinin başarılı bir şekilde gerçekleştirilmesi için derin öğrenme

mimari modellerine literatürdeki çalışmalarda başvurulmuştur. Bunun sebebi gelişen teknolojik yapıda yapay zekânın azımsanamaz bir yer alması ve bu gibi durumların ortadan kaldırılması için yüksek başarımda sonuçlar elde edildiğinin kanıtlanmış olmasıdır.

Literatürde yer alan çalışmalarda CNN mimarisi kullanılarak görüntü veri kümeleri üzerinden çıkarsaması yapılan basit sayısal değerlerle ifade edilen öznitelikler, görüntü hakkında araştırmacılara temel bilgiler sunarken, soyutlama yoluyla oluşturulan derin öznitelikler ise görüntü hakkında anlamsal olarak zenginleştirilmiş bilgi barındırmaktadır. Anlamsal boşluk olarak adlandırılan problemi gidermenin bir yolu ontoloji kavramından geçmektedir. Görüntü ontolojileri, insanların görüntüler aracılığıyla algıladıkları üst düzey kavramları tanımlamak amacıyla kullanılır. Anlamsal ağ kavramı geliştikçe ontolojiler, bilgiyi makinelerin işlem süzgecinden geçirerek sonuç

çıkartabileceği bir biçime dönüştürmek ve bu belirtilen bilgi kümelerinin aralarında anlam bakımından tutarlı ilişkililik kurmak için kullanılmaktadır. Ontolojilerde tanımlanan kavramların bazıları kullanılarak teknolojik gelişim için bir engel olan anlamsal boşluk problemlerinin giderilmesinde etkin ve başarıyı artıracak çıktılar sağladığı gösterilmiştir. Bu şekilde nesne kümelerinin algılanması ve algılanan bu nesnelerin tanımlanmasında farklı bilgi alanları kullanılmıştır. WordNet [27] bunlara en güzel örneklerden biridir. Sözlük veri kümesi şeklinde günümüzdeki çalışmalarda kullanılmaktadır. Literatürde bu alana özgü daha önceki çalışmalarda da incelemeler yapılmıştır [36].

Teşekkür

Bu çalışmayı inceleyen Veri Bilimi dergisi hakemlerine destekleri için teşekkür ederiz.

Kaynaklar

- [1] Alpaydın, E. (2004). Introduction to machine learning. MIT Press.
- [2] Aslandogan, Y. A., & Yu, C. T. (1999). Techniques and systems for image and video retrieval. In IEEE Transactions on Knowledge and Data Engineering (Vol. 11, Issue 1, pp. 56-63). <https://doi.org/10.1109/69.755615>
- [3] Alkhwilani, M., Elmogy, M., & El Bakry, H. (2015). Text-based, content-based, and semantic-based image retrievals: A survey. In International Journal of Computer and Information Technology (ISSN: 2279-0764) (Vol. 4, Issue 01).
- [4] Ngo, T. G., Ngo, Q. T., & Nguyen, D. D. (2016). Image Retrieval with relevance feedback using SVM active learning. International Journal of Electrical and Computer Engineering, 6(6), 3238-3246. <https://doi.org/10.11591/ijece.v6i6.11631>
- [5] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. In IEEE. <https://doi.org/10.1109/CVPR.2015.7298965>
- [6] El Shaer, Mohamed & Wisdom, Scott & Mishra, Taniya. (2019). Transfer Learning From Sound Representations For Anger Detection in Speech. Erişim Adresi: https://www.researchgate.net/publication/330923908_Transfer_Learning_From_Sound_Representations_For_Anger_Detection_in_Speech
- [7] Tzelepi, M., & Tefas, A. (2018). Deep convolutional learning for Content Based Image Retrieval. In Neurocomputing (Vol. 275, pp. 2467-2478). <https://doi.org/10.1016/j.neucom.2017.11.022>
- [8] Oxford 5K Veri Kümesi. Erişim Adresi (23.06.2023): <https://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>
- [9] Alzu'bi, A., Amira, A., & Ramzan, N. (2017). Content-based image retrieval with compact deep convolutional features. In Neurocomputing (Vol. 249, pp. 95-105). <https://doi.org/10.1016/j.neucom.2017.03.072>
- [10] Noh, H., Hong, S., & Han, B. (2015). Learning Deconvolution Network for Semantic Segmentation (Vol. 1). In: ICCV.. <https://doi.org/10.1109/ICCV.2015.178>
- [11] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481-2495. <https://doi.org/10.1109/TPAMI.2016.26444615>
- [12] De Geus, D., Meletis, P., & Dubbelman, G. (2020). Fast panoptic segmentation network. IEEE Robotics and Automation Letters, 5(2), 1742-1749. <https://doi.org/10.1109/LRA.2020.2969919>
- [13] Suhasini, P. S., Krishna, K. S. & Krishna, I. V. (2008). Graph Based Segmentation in Content Based Image Retrieval. Journal of Computer Science, 4(8), 699-705. <https://doi.org/10.3844/jcssp.2008.699.705>
- [14] Ozden M., Polat E. (2007) A color image segmentation approach for content-based image retrieval. Pattern Recognition 40(4):1318-1325. <https://doi.org/10.1016/j.patcog.2006.08.013>
- [15] Rizwan I Haque, I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. In Informatics in Medicine Unlocked (Vol. 18). <https://doi.org/10.1016/j.imu.2020.100297>
- [16] Ma, H., Zhu, J., Lyu, M. R. T., & King, I. (2010). Bridging the semantic gap between image contents and tags. IEEE Transactions on Multimedia, 12(5), 462-473. <https://doi.org/10.1109/TMM.2010.2051360>
- [17] Pang, Y., Li, Y., Shen, J., & Shao, L. (2019). Towards bridging semantic gap to improve semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision, 2019-Octob(Iccv), 4229-4238. <https://doi.org/10.1109/ICCV.2019.00433>
- [18] Philbin, J., Chum, O., Isard, M., & Sivic, J., (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR. 1-8.
- [19] Zhao, R., & Grosky, W. I. (2002). Narrowing the semantic gap - Improved text-based web document retrieval using visual features. IEEE Transactions on Multimedia, 4(2), 189-200. <https://doi.org/10.1109/TMM.2002.1017733>
- [20] Wu, L., Hua, X.-S., Yu, N., Ma W.-Y., & Li, S. (2008). Flickr distance, In: Proc. MM'08, 31-40, WordNet Veri Kümesi. Erişim Adresi (23.06.2023): <https://wordnet.princeton.edu>
- [21] Auchard, E., (2007). Flickr to map the world's latest photo hotspots, Reuters, CNet. Erişim Adresi (23.06.2023): <https://www.cnet.com/tech/tech-industry/flickr-to-map-the-worlds-latest-photo-hot-spots/>
- [22] Everingham, M., Eslami, S. M. A. Van Gool, L. J., Williams, C. K. I. Winn J. M. & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective, IJCV, 2015.
- [23] Brostow, G. J. Shotton, J. Fauqueur J. & Cipolla, R. (2008). Segmentation and recognition using structure from motion point clouds, In: Proc. ECCV.
- [24] Scopus Web Sitesi. Erişim Adresi (23.06.2023): <https://www.scopus.com/home.uri>
- [25] Web of Science Web Sitesi. Erişim Adresi (23.06.2023): <https://access.clarivate.com/>
- [26] IEEE Web Sitesi. Erişim Adresi (23.06.2023): <https://ieeexplore.ieee.org/Xplore/home.jsp>
- [27] WordNet Veri Kümesi. Erişim Adresi (23.06.2023): <https://wordnet.princeton.edu>

- [28] Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., & Jeon, G. (2018). Content Based Image Retrieval by Using Color Descriptor and Discrete Wavelet Transform. *Journal of Medical Systems*, 42(3). <https://doi.org/10.1007/s10916-017-0880-7>
- [29] Yenigalla S.C., Rao K.S., Ngangbam P.S.. (2023) Implementation of Content-Based Image Retrieval Using Artificial Neural Networks. *Engineering Proceedings*; 34(1):25. <https://doi.org/10.3390/HMAM2-14161>
- [30] Sikandar S., Mahum R., Alsalman A. (2023) A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion. *Applied Sciences*; 13(7):4581. <https://doi.org/10.3390/app13074581>
- [31] Song, K., Li, F., Long, F., Wang, J., & Ling, Q. (2018). Discriminative Deep Feature Learning for Semantic-Based Image Retrieval. *IEEE Access*, 6, 44268– 44280. <https://doi.org/10.1109/ACCESS.2018.2862464>
- [32] Wu, Q. (2020). Image retrieval method based on deep learning semantic feature extraction and regularization softmax. *Multimedia Tools and Applications*, 79(13–14), 9419–9433. <https://doi.org/10.1007/s11042-019-7605-5>
- [33] Bouchakwa, M., Ayadi, Y., & Amous, I. (2020). Multi-level diversification approach of semantic-based image retrieval results. *Progress in Artificial Intelligence*, 9(1), 1–30. <https://doi.org/10.1007/s13748-019-00195-x>
- [34] Li, Y., Wang, Y., & Huang, X. (2007). A relation-based search engine in Semantic Web. *IEEE Transactions on Knowledge and Data Engineering*, 19(2), 273–281. <https://doi.org/10.1109/TKDE.2007.18>
- [35] Minu, R. I., & Thyagarajan, K. K. (2014). Semantic rule based image visual feature ontology creation. *International Journal of Automation and Computing*, 11(5), 489–499. <https://doi.org/10.1007/s11633-014-0832-3>
- [36] Gasi, A., Ensari, T. ve Dagtekin, M. (2021). Anlamsal tabanlı görüntü erişimi üzerine bir derleme. *Acta Infologica*, 5(2), 445-457. <https://doi.org/10.26650/acin.835241>