ACTA INFOLOGICA 2021;5(2):445-457
dergipark.org.tr/acin

DOI: 10.26650/acin.835241

REVIEW ARTICLE

Anlamsal Tabanlı Görüntü Erişimi Üzerine Bir Derleme

Semantic Based Image Retrieval-A Survey

Akif Gaşi[1], Tolga Ensari[2], Mustafa Dağtekin[3]

[1]İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye
[2](Asst. Prof.), Arkansas Tech University, Department of Computer and Information Science, Arkansas, USA
[3](Dr. Öğr. Üyesi), İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

ORCID: A.G.0000-0001-8049-1273;
T.E. 0000-0003-0896-3058;
M.D. 0000-0002-0797-9392

Corresponding author:
Akif GAŞİ
İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye
E-mail address: akif.gasi@ogr.iu.edu.tr

Submitted: 11.12.2020
Revision Requested: 27.05.2021
Last Revision Received:07.06.2021
Accepted: 09.06.2021
Published Online: 10.09.2021

Citation: 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

ÖZ

Bilgisayar teknolojisinin hızlı gelişmesi sonucunda çok büyük miktarlarda verilerle çalışmak olağan bir durum haline gelmiştir. Bu verilerin işlenmesi, verilerden anlamlı bilgiler çıkarılması ve kararlar alınması için makine öğrenmesine ihtiyaç duyulmaktadır. Görüntü içerisinden nesnelerin algılanmasına yönelik son zamanlarda yapılan çalışmalar özellikle Anlamsal Tabanlı Görüntü Erişimi alanına doğru yönelmektedir. Bu alanda yapılan çalışmalar ile Anlamsal Boşluk olarak adlandırılan ve görüntülerden makineler tarafından çıkarılan düşük düzeydeki renk, şekil, doku (color, shape, texture) özellikleri ile insanlar tarafından resimlerden algılanan ve yüksek düzey olarak ifade edilen kavramlar arasındaki uyuşmazlıkların giderilmesine yoğunlaşmaktadır. Bu amaçla, belirli bir bilgi alanına (domain) ait kavramların özelliklerini ve aralarındaki ilişkileri göstermek için iyi tanımlanmış ontolojiler oluşturulmakta ve arama işlemi bu yönde ilerlemektedir. Ontolojiler kullanılarak bilgiler bilgisayarların işleyebileceği biçime dönüştürülmekte ve bilgiler arasında anlamlı ilişkiler oluşturulabilmektedir. Bu çalışmada Anlamsal Tabanlı Görüntü Erişimi (Semantic Based Image Retrieval - SBIR) üzerine bir derleme yapılmıştır. SBIR ile amaç İçerik Tabanlı Görüntü Erişimi (Content Based Image Retrieval - CBIR) ile yapılan arama işlemlerinde karşılaşılan ve Anlamsal Boşluk (Semantic Gap) olarak ifade edilen darboğazın aşılmasıdır.

Yapılan çalışmalarda Ontoloji (Ontology) kavramının kullanılmasıyla problemin çözümünde önemli bir gelişme yaşandığı gözlemlenmiştir.

Anahtar kelimeler: Görüntü Alma, Anlamsal Boşluk, Derin Öğrenme, Semantik Görüntü Alma

ABSTRACT

As a result of rapid technological development, operating with massive data has become a common situation. There is a need for machine learning to process these data and extract meaningful information, and make a decision from them. Current studies related to identifying objects from the image are driven to Semantic-Based Image Retrieval. The studies done in this field aim to dismiss the discrepancies among the low-level color, shape, texture characteristics and picture recognition by people that are extracted from images by machines known as the Semantic Gap, that are signified as high-level concepts. Therefore, definite ontologies are created to determine characteristics of the concept of a particular domain and show the relationship between them by advancing the research on this area. Through ontologies, information is transformed into a structure so computers can process and create a meaningful relationship between information. In this study, a compilation on Semantic-Based Image Retrieval – SBIR is done. SBIR aims to overcome the bottleneck faced in the search operations created by Content-Based Image Retrieval (CBIR) and shown as a Semantic Gap.

In the studies done, significant progress in problem-solving through the use of the Ontology concept is observed.

Keywords: Image Retrieval, Semantic Gap, Deep Learning, Semantic Based Image Retrieval

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

1. GİRİŞ

Bilgisayar teknolojisinin hızlı gelişmesi sonucunda çok büyük miktarlarda verilerle çalışmak olağan bir durum haline gelmiştir. Bu verilerin işlenmesi, verilerden anlamlı bilgiler çıkarılması ve kararlar alınması için makine öğrenmesine ihtiyaç duyulmaktadır. Makine öğrenmesi bilgisayar bilimleri, istatistik, eğitim, sağlık, sanayi/üretim, finans, reklam gibi birçok alanda kullanılmaktadır. Makine öğrenmesi bilgisayarların deneyimler yoluyla problem çözme yaklaşımlarının nasıl iyileştirilebileceği ya da geliştirebileceği sorusunu ele almaktadır. Makinelerin öğrenme yeteneğinin iyileştirilmesi veya geliştirilmesi temel amaçtır (Alpaydın, 2004).

Görüntü içerisinden nesnelerin algılanmasına yönelik son zamanlarda yapılan çalışmalar özellikle Anlamsal Tabanlı Görüntü Erişimi alanına doğru yönelmektedir. Bu alanda yapılan çalışmalar ile Anlamsal Boşluk olarak adlandırılan ve görüntülerden makineler tarafından çıkarılan düşük düzeydeki renk, şekil, doku (color, shape, texture) özellikleri ile insanlar tarafından resimlerden algılanan ve yüksek düzey olarak ifade edilen kavramlar arasındaki uyuşmazlıkların giderilmesine yoğunlaşmaktadır. Kavramların bilgisayarlar tarafından anlaşılabilir ve işlenebilir hale getirilmesi amacıyla Ontolojiler kullanılmaktadır. Ontoloji, varlığın doğasıyla ve varlıklar arasındaki ilişkiler ile ilgilenen metafizik dalıdır. Bu amaçla, belirli bir bilgi alanına (domain) ait kavramların özelliklerini ve aralarındaki ilişkileri göstermek için iyi tanımlanmış ontolojiler oluşturulmakta ve arama işlemi bu yönde ilerlemektedir.

Bu çalışmanın içeriği şu bölümlerden oluşmaktadır: Görüntü Erişimi işleminde kullanılan temel konular üzerinde açıklamalar 2. bölümde yapılmıştır. Anlamsal Boşluk kavramı ile ilgili yapılmış çalışmalara 3. bölümde yer verilmiştir. Anlamsal Tabanlı Görüntü Erişimi alanında yapılmış temel çalışmalar 4. bölümde açıklanmıştır. Çalışmanın sonuç kısmına ise 5. bölümde yer verilmiştir.

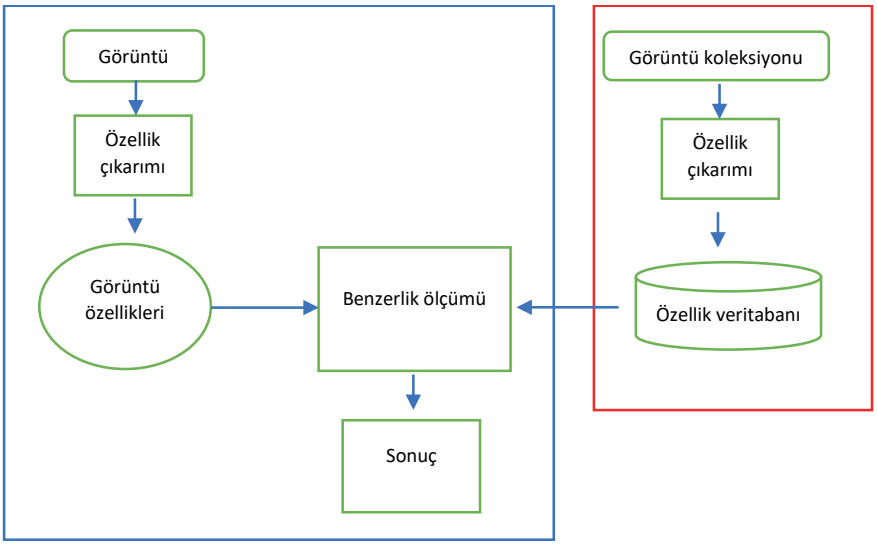
2. GÖRÜNTÜ ERİŞİMİ (IMAGE RETRIEVAL)

Görüntü erişimi, kullanıcıların sorgu amaçlı kullandığı görüntü ile benzer olan resimlerin bir veri tabanında aranması işlemi ile ilgilidir (Aslandogan ve Yu, 1999). Görüntü erişimi işlemi genelde üç farklı yol ile yapılmaktadır. Birinci yol Yazı Tabanlı Görüntü Erişimi (Text Based Image Retrieval - TBIR), ikinci yol İçerik Tabanlı Görüntü Erişimi, üçüncü yol ise Anlamsal Tabanlı Görüntü Erişimi olarak ifade edilir.

Birinci yoldan yapılan arama işlemleri, resim içeriği hakkında bilgi veren anahtar kelime (keyword) temelli yapıldığından genelde aranılan kelime ile hiçbir bağlantısı olmayan resimler de sonuç olarak getirilmektedir. TBIR işleminden önce resimlerin içeriği hakkında bilgi veren bir yazı ile manuel ya da otomatik olarak açıklama eklenmesi gerekmektedir. Manuel olarak yapılan açıklama ekleme işlemi zaman alıcı bir süreçtir ve yanlışlara neden olmaktadır. Buna alternatif olarak önerilen bir çalışmada (Alkhawlani, vd., 2015), Otomatik Görüntü Altyazılama (Automatic Image Captioning) yöntemi ile resimlerin içeriğine göre uygun bir açıklama eklenmektedir.

CBIR (Smeulders, vd., 2000) diye tabir edilen yöntemde, görüntüden algılanan nesnelere ait çıkarılan renk, şekil, doku özellikleri ile daha önceden veri tabanında saklanan görüntü özelliklerinin benzerlik durumlarına göre arama yapılmaktadır. Görüntü Erişimi ve özellikle İçerik Tabanlı Görüntü Erişimi alanındaki gelişmeler ve bu alanda geliştirilen Görüntü Erişimi sistemlerinin incelendiği bir çalışmada (Y., T.S., ve S-F. 1999), sistem başarımını arttırmak için indeksleme (indexing) ve boyut indirgeme (dimensionality reduction) işlemlerinin öneminden bahsedilmiş, farklı disiplinlerin bir arada kullanılmasıyla ve insan-bilgisayar etkileşiminin geliştirilmesiyle gelecekte Görüntü Erişimi sistemlerinin başarımının arttırılacağından bahsedilniştir. Yapılan diğer bir çalışmada ise (Alzu'bi, vd., 2015) CBIR yönteminde kullanılan Görüntü Bölütleme (Image Segmentation), Özellik Çıkarımı (Feature Extraction), bu alanda kullanılan Makine Öğrenmesi yöntemleri, sonuçların değerlendirilmesinde kullanılan Başarım Ölçümü (Performance Evaluation), arama işleminde kullanılan Görüntü Veri Kümeleri (Image Dataset) hakkında detaylı bir derleme yapılmış ve sisteminin genel çerçevesi Şekil 1'de gösterilmiştir.

Araştırmamızın ana konusunu oluşturan Anlamsal Tabanlı Görüntü Erişimi yöntemi ile anlamsal benzer içeriğe sahip görüntülerin aranmasına yönelik çalışmalar yapılmaktadır. Anlamsal Tabanlı Görüntü Erişimi yöntemi hakkında detaylı bilgiye 4. bölümde yer verilmiştir.



Şekil 1. İçerik Tabanlı Görüntü Erişimi sistem mimarisi (Alzu'bi, Amira, ve Ramzan, 2015)

Görüntü erişimi işleminin iyileştirilmesi için yapılan bir çalışmada (Ngo, vd., 2016), Destek Vektör Makineleri (Support Vector Machine - SVM) öğrenme yöntemi ve İlgililik Geribildirimi (Relevance Feedback - RF) birlikte kullanılmıştır. SVM ile veri tabanından arama işleminde kullanılan görüntüye benzer resimler alınmıştır. İlgililik Geribildirimi ile kullanıcı ve sistem arasında sağlanan etkileşimle, Anlamsal Boşluk kavramını en aza indirgemek için kullanıcının veri tabanından alınan sonuçlar üzerinde verdiği kararlara göre algoritmanın eğitilmesi yapılmaktadır.

Görüntü erişimi işleminde alınan sonuçların iyileştirilmesi amacıyla, resimlerin gösterimi (Image Representation) için yürütülen çalışmalar önemli bir yer kaplamaktadır. Yapılan başka bir çalışmada (Wang, vd., 2020), Aggregating Convolutional Kernel - ACK diye isimlendirilen yöntem ile resimlerin gösterimi için yeni bir yaklaşım sunulmuştur. Önerilen yöntemde Evrişimli Sinir Ağı (Convolutional Neural Network - CNN) çekirdekleri özellik çıkarımı için kullanılmaktadır. CNN ile görüntüye ait düşük düzeydeki renk, şekil ve doku gibi özelliklerden başlayarak seviyeli bir şekilde yüksek düzey diye ifade edilen ve insanlar tarafından algılanan anlamsal kavramlara ait özelliklerin çıkarımı aşamalı bir şekilde gerçekleştirilmektedir. Önerilen yöntemde, resimlerin parlaklık özelliği kullanılarak Evrişimli Sinir Ağına ait belirli bir çekirdeğin etkinleştirilmesi gerçekleştirilmiştir. Parlaklık yoğunluğuna göre her seviyede uygun çekirdeğin seçimi yapılarak bu çekirdekler resimlerin gösterimi için kullanılmıştır.

İnternet teknolojisinin hızlı gelişmesiyle birlikte internet ortamında kullanılan resimlerin boyutunda da ciddi bir artış yaşanmıştır. Büyük boyuttaki multimedya verilerinin aranması işleminde, insanın görme duyusu temel alınarak geliştirilen bir algoritmada (Zhu, 2020), önce renk ve doku özellikleri kullanılarak aday görüntü bölgeleri oluşturulmuş, sonrasında ise belirlenen bölgelerde derin öğrenme yaklaşımı ile özellik çıkarımları yapılmış ve geliştirilen bir Renk Algılama Fonksiyonu (Color Attention Function) ile resimlerin benzerlik ölçümü yapılarak, sıralama sonucuna göre üst sıralarda yer alan resimler sonuç olarak gösterilmiştir.

Görüntü Alma başlığı altında incelenen makalelerde kullanılan Yöntem, Veri kümesi, Algoritma, Başarım ölçeği ve Sonuçlara Tablo 1'de yer verilmiştir.

Tablo 1

Görüntü Erişimi başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

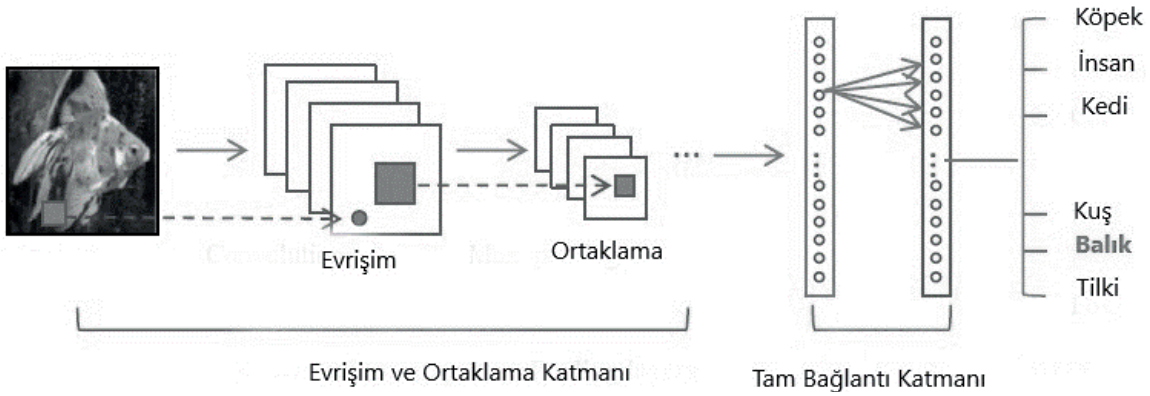
Makale	Yöntem	Veri kümesi	Algoritma	Başarım ölçeği	Sonuç
(Ngo, Ngo, ve Nguyen 2016)	Destek Vektör Makineleri (Support Vector Machines - SVM)	Corel Gallery	Renk anları (Color moments), Piramidal Dalgacık Dönüşümü (Pyramidal Wavelet Transform-PWT), Kenar Yönü Histogramı (Edge Direction Histogram-EDH)	Ortalama hassasiyet (Average precision-AP)	0.85
(Q. Wang, vd., 2020)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	Oxford Paris Holidays	Evrişim Çekirdeklerini Birleştirme (Aggregating Convolution Kernels-ACK)	Ayrık Sıralı Diziler (Discrete Ordered Sequences)	64.98 76.26 84.72
(Zhu 2020)	Derin Sinir Ağları (Deep Neural Network)	Cifar Caltech	İkili Normlu Eğim (Binarized Normed Gradient - BING)	Minkowski	0.9473 0.9462

2.1 İçerik Tabanlı Görüntü Erişimi

CBIR ile arama işlemi, sorguda kullanılan görüntüye ait çıkarılan renk, şekil ve doku özellikleri ile veri tabanında saklanan özelliklerin benzerlik durumlarına göre yapılmaktadır. Görüntüden çıkarılan renk özelliği analizi temel alınarak yapılan bir çalışmada (Ashraf, vd., 2018), YCbCr ve RGB renk uzayları görüntüye ait renk özelliklerinin çıkarımında kullanılmıştır. Renklerin parlaklık özelliği temel alınarak ilk önce RGB renk uzayından YCbCr renk uzayına dönüşüm yapılmıştır. Sonra, sorguda kullanılan görüntü Y, Cb, Cr renk bileşenlerine ayrılmıştır. Y kısmına Canny Kenar Tespiti (Canny Edge Detector) işlemi uygulanarak kenarlar belirlenmiştir. İşlemden sonra, görüntü tekrar RGB renk uzayına dönüştürülür. RGB dönüştürme işleminden sonra görüntü tekrar R, G, B bileşenlerine ayrılır ve her bileşen için çizit (histogram) hesaplanır. Son adımda ayrıştırılan her histogram için Ayrık Dalgacık Dönüşümü (Discrete Wavelet Transform - DWT) işlemi, görüntünün mekansal alandan (spatial domain) frekans alanına (frequency domain) dönüştürülmesi için kullanılmış ve özellik vektörü hesaplanmıştır. Özellik vektörü hesaplandıktan sonra Yapay Sinir Ağları (Artificial Neural Network - ANN) ile sınıflar arasındaki anlamsal ilişkiler belirlenmiştir. Evrimsel programlama (Evolutionary Programming) yaklaşımının kullanıldığı bir çalışmada (Jadhav ve Patil 2012)geliştirilen CBIR sisteminde, görüntülerden çıkarılan renk, şekil ve doku özelliklerinin benzerlik durumlarına göre yapılan görüntü erişimi denemelerinde, Precision 0.31 ve Recall 0.70 başarım değerleri elde edilmiştir.

Derin Öğrenme (Deep Learning) mimarilerinin gelişmesiyle birlikte, makine öğrenmesi alanındaki çalışmalar büyük ivme kazanmıştır. Derin Öğrenme mimarileri ile yüksek seviye diye ifade edilen ve insanlar tarafından görüntüden algılanan kavramların bilgisayar tarafından öğrenilmesi amaçlanmaktadır. Şekil 2'de gösterilen Evrişimli Sinir Ağları mimarileri, kavramlara ait özelliklerin çıkarılması amacıyla kullanılmaktadır.

Yapılan bir çalışmada (Guo ve diğerleri., 2016), Derin Öğrenme yöntemlerinin görüntü işleme alanında kullanımı ile ilgili detaylı açıklamalara yer verilmiştir.



Şekil 2. Evrişimli Sinir Ağları mimarisi (Guo, vd., 2016)

Deep CNN mimarisinin yeniden eğitilmesi temel alınarak geliştirilen bir çalışmada (Tzelepi ve Tefas, 2018), veri kümesi hakkında herhangi bir bilgi olmaması durumunda Tam Denetimsiz Yeniden Eğitim (Fully Unsupervised Retraining) yaklaşımı kullanılmış, veriler hakkında bir sınıf etiketi (class label) bulunması durumunda var olan sınıf etiketi kullanılmış, kullanıcı etkileşimi olması durumunda İlgililik Geribildirimi ile sonuçlar üzerinde kullanıcılar tarafından verilen kararlar kullanlarak algoritmanın yeniden eğitilmesi gerçekleştirilmiştir.

CNN kullanılarak yapılan diğer bir çalışmada (Alzu'bi, vd., 2017), paralel iki farklı CNN mimarisi kullanılarak özellik çıkarımı yapılmış ve Oxford 5K veri kümesi üzerinde %95 başarı değeri elde edildiği belirtilmiştir.

İçerik Tabanlı Görüntü Erişimi başlığı altında incelenen makalelerde kullanılan Yöntem, Veri kümesi, Algoritma, Başarım ölçeği ve Sonuçlara Tablo 2'de yer verilmiştir.

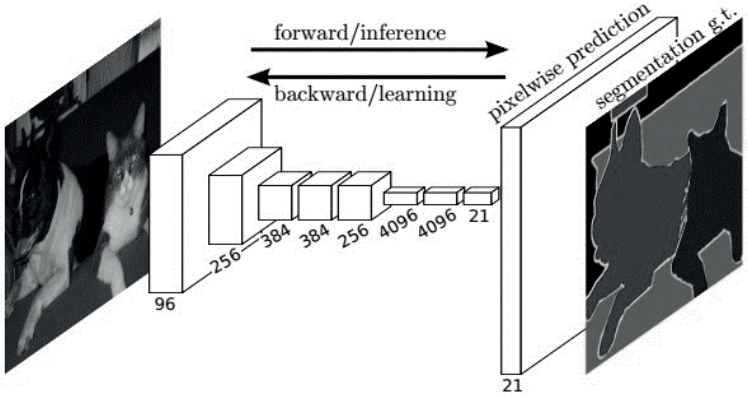
Tablo 2

İçerik Tabanlı Görüntü Erişimi başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

Makale	Yöntem	Veri kümesi	Algoritma	Başarım ölçeği	Sonuç
(Ashraf, vd., 2018)	Yapay Sinir Ağları (Artificial Neural Network)	Corel	Ayrık Dalgacık Dönüşümü (Discrete wavelet transform - DWT)	Hassasiyet (Precision), Geri çağırma (Recall)	0.51 0.30
(Tzelepi ve Tefas 2018)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	Paris 6K UKBench	Uyarlanabilir an tahmini algoritması (Adaptive moment estimation algorithm - ADAM)	Ortalama hassasiyet (Mean Average Precision - mAP)	Denetimli - Supervised (9.9859, 3.9710) Denetimsiz - Unsupervised (0.8347, 3.8323)
(Alzu'bi, Amira, ve Ramzan 2017)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	Oxford 5K Holidays	İki çizgili model (Bilinear model)	Ortalama hassasiyet (Mean Average Prcision)	%95.7 %95.1

2.2 Bölütleme (Segmentation)

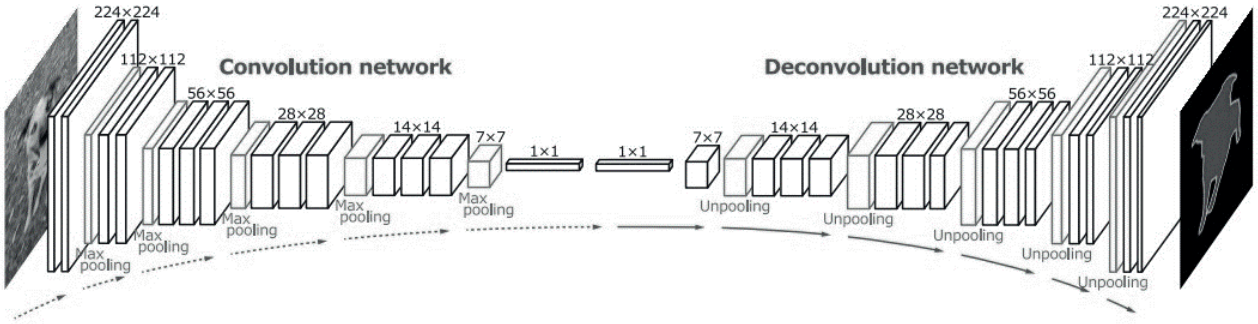
Bölütleme işlemi, görüntüye ait her piksele bir sınıf etiketinin atanmasını temel alan bir modeldir. Görüntüye atanan piksel temelli anlamsal bilgiler, nesnelerin algılanması ve kararlar alınması için kullanılmaktadır. CNN mimarisi Bölütleme işleminde kullanılan ve etkili sonuçlar veren bir modeldir. Yapılan bir çalışmada (Long, vd., 2015) önerilen Tam Evrişimli Ağ (Fully Convolutional Network - FCN) modeli, sınıflandırma algoritmalarının (AlexNET, VGGNet, GoogleNet) adaptasyonu ile piksel temelli eğitildiğinde Şekil 3'te gösterildiği gibi bölütleme işleminde etkili sonuçlar verdiği gösterilmiştir.



Şekil 3. Tam Evrişimli Ağ ile bölütleme işlemi (Long, Shelhamer, ve Darrell 2015)

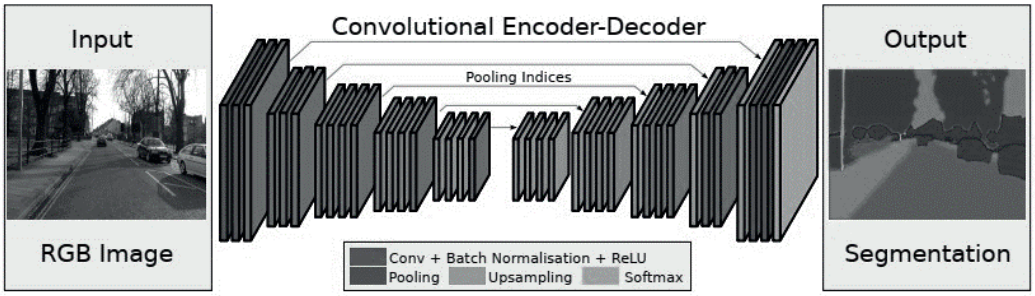
FCN bölütleme işlemini yaparken görüntüde sabit boyuttaki alanlar (fixed-size fields) kullanmakta ve bu durum belirlenen alandan daha büyük ya da daha küçük olan nesnelerin parçalanması ya da herhangi bir sınıf etiketinin eklenmemesine neden olmaktadır. Genelde sınıf etiketinin eklenmesi işlemi, büyük boyuttaki nesneler için yapılmaktadır. Küçük boyuttaki nesneler ise göz ardı edilmekte ya da görüntüde arka plan olarak sınıflandırılmaktadır.

Yapılan diğer bir çalışmada önerilen Evrişimli - Ters Evrişimli (Convolutional - Deconvolutional) mimarisinin kullanıldığı modelde (Noh, vd., 2015), FCN ile karşılaşılan eksikliklerin giderilmesi amacıyla örnek bölütleme (instance segmentaion) yapılmış, Evrişimli Ağ (Convolutional Network) ile özelliklerin çıkarımı ve Ters Evrişimli Ağ (Deconvolutional Network) ile çıkarılan özellikler kullanılarak nesne bölütleme işlemi yapılmıştır. Şekil 4'te, önerilen Evrişimli - Ters Evrişimli Ağ mimarisi gösterilmiştir.



Şekil 4. Evrişimli - Ters Evrişimli Ağ mimarisi (Noh, Hong, ve Han 2015)

Önerilen diğer bir çalışmada (Badrinarayanan, vd., 2017) geliştirilen piksel temelli anlamsal bölütleme (semantic segmentation) yapan SegNet mimarisi (yol ve kapalı alan görüntülerinin algılanması) kodlayıcı (encoder), kod çözücü (decoder) ve piksel temelli sınıflandırma katmanlarından oluşmaktadır. Temel aldığı VGG-16 mimarisinin Tam Bağlantılı Katmanının (Fully Connected Layer) çıkarılmasıyla ve bunun yerine piksel temelli sınıflandırma katmanının eklenmesiyle Şekil 5'te gösterildiği gibi daha küçük ve daha hızlı bir mimari elde edilmiştir.



Şekil 5. SegNet mimarisi (Badrinarayanan, vd., 2017)

Panoptik bölütleme Ağı (Panoptic Segmentation Network), Anlamsal Bölütleme Ağı (Semantic Segmentation Network) ve Örnek Bölütleme Ağı (Instance Segmentation Network) modellerinin birleşimidir. Panoptic segmentation ile görüntüye ait her piksele bir sınıf etiketi eklenmektedir. Yapılan bir çalışmada (De Geus, vd., 2020), önerilen Hızlı Panoptik Bölütleme Ağı (Fast Panoptic Segmentation Network - FPSNet) ile görüntüde yer alan nesneler Varlık Sınıfı (Things Classes - sayılabilir nesneler: insanlar, arabalar) ve Soyut Sınıf (Stuff Classes - sayılamayan nesneler: gökyüzü, su) gibi farklı sınıflara ayrılmakta ve bölütleme işlemi gerçekleştirilmektedir. Şekil 6'da, Panoptik bölütleme işleminde girdi olarak kullanılan görüntü ve panoptic bölütleme yapıldıktan sonra elde edilen sonuç gösterilmiştir.



Şekil 6. Panoptik bölütleme (Panoptic Segmentation) (De Geus, vd., 2020)

Derin öğrenme mimarileri ve özellikle CNN mimarisi, sınıflandırma dışında resim bölütleme işlemlerinde de sıklıkla kullanılmaktadır. Medikal görüntülemede derin öğrenme yaklaşımı kullanılarak resim bölütleme işlemi için CNN ve FCN mimarileri etkili sonuçlar vermektedir. Yapılan bir çalışmada (Haque ve Neubert, 2020), medikal görüntülemede resim bölütleme işleminde kullanılan derin öğrenme teknikleri bir tablo yardımıyla özetlenmiştir. Tabloda makalelerde kullanılan yöntemlere, açıklamalara, performans ölçeklerine ve sonuçlara yer verilmiştir.

Görüntüden bölütleme amaçlı toplanan veriler birden fazla kaynaktan elde edilebilir. Bu durum, Çoklu Kip (Multiple Modality) olarak isimlendirilir. Çoklu kip, her bir algılayıcının aynı sahne hakkında bilgi toplaması işlemi olarak tanımlanır. Çoklu kip yaklaşımında birden fazla kaynaktan alınan ve birbirini tamamlayan bilgiler (Complementary Information) mevcuttur. Birden fazla kaynaktan alınan bu bilgiler anlamsal algılama ve karar verme işlemlerinde kullanılmaktadır. Yapılan bir çalışmada (Zhang, vd., 2020), derin çoklu kip yöntemleri hakkında bir inceleme yayınlanmıştır. Bu alanda yapılan ilk çalışma olması açısından yapılan çalışmalar, hiyerarşik bir yaklaşımla sıralanmış ve bir tablo yardımıyla özetlenmiştir.

Bölütleme başlığı altında incelenen makalelerde kullanılan Yöntem, Veri kümesi, Başarım ölçeği ve Sonuçlara Tablo 3'te yer verilmiştir.

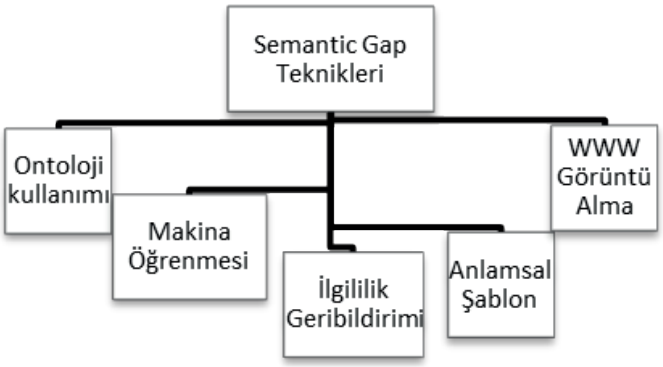
Tablo 3

Bölütleme başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

Makale	Yöntem	Veri kümesi	Başarım ölçeği	Sonuç
(Long, Shelhamer, ve Darrell 2015)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	PASCAL VOC NYUDv2 SIFT Flow	Ortalam birleşim (Mean IU),	62.6 34.0 39.5
(Noh, Hong, ve Han 2015)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	PASCAL VOC 2012	Birleşim üzerinde kesişim (Intersection over Union-IoU)	72.5
(Badrinarayanan, Kendall, ve Cipolla 2017)	Tam Evrişimli Ağ (Fully Convolutional Network - FCN)	PASCAL VOC 2012	Birleşim üzerinde ortalama kesişim (Mean Intersection over Union)	60.10
(De Geus, Meletis, ve Dubbelman 2020)	Hızlı Panoptik Bölütleme Ağı (Fast Panoptic Segmentation Network - FPSNet)	PASCAL VOC 2012 Cityscapes	Panoptik kalite ölçeği (Panoptic Quality Metric)	57.8 55.1

3. ANLAMSAL BOŞLUK

Anlamsal Boşluk kavramı CBIR işleminde karşılaşılan temel sorun olup resimlerden çıkarılan düşük düzeydeki özellikler renk, şekil, doku ve insanların görüntülerden algıladığı yüksek düzeydeki kavramlar arasındaki uyuşmazlık olarak tanımlanır. Yapılan bir çalışmada (Liu, vd., 2007) karşılaşılan uyuşmazlıkları gidermek için Şekil 6'de gösterilen: 1) İnsanlar tarafından algılanan yüksek düzeydeki kavramların tanımlanmasında Ontolojilerin kullanılması; 2) Sorgu ifadesi ile düşük düzeydeki özelliklerin ilişkilendirilmesi amacıyla Makine Öğrenmesi yöntemlerinin kullanılması; 3) İlgililik Geribildirimi ile kullanıcının elde edilen sonuçlar üzerinde verdiği kararlar ile sistemin eğitilmesi 4) Görüntü erişimi işleminde düşük düzeydeki özellikler ile yüksek düzeydeki kavramlar arasında bağlantının oluşturulması için Anlamsal Şablon (Semantic Template) oluşturulması; 5) Web ortamında görüntü alma amacıyla HTML yazı ve resimlerin birleştirilmesi gibi tekniklerin kullanılması önerilmektedir.



Şekil 6. Bölütleme başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

Web belgelerinde var olan anlamsal boşluk sorununun giderilmesi için yazı ve resim özellikleri web belgelerinin aranmasında kullanılmaktadır. Yapılan bir çalışmada (Zhao ve Grosky, 2002), web belgeleri üzerinde karşılaşılan anlamsal boşluk sorununun çözümünde Saklı Anlamsal İndeksleme (Latent Semantic Indexing - LSI) yöntemi ile semantik açıdan benzer kelimeler aynı küme içinde toplanarak, arama işleminde kullanılan anahtar kelimeyi içeren web belgelerinin aranmasında kullanılmıştır. Haber servis siteleri üzerinde yapılan arama işlemlerinde, LSI ile geliştirilen yaklaşımda etkili sonuçlar alındığı gösterilmiştir.

Bölütleme amacıyla resimlerden çıkarılan sığ katmanlardaki özellikler görüntüye ait detaylı bilgileri içerirken (kenarlar, çizgiler, köşeler), derin katmanlardaki özellikler ise görüntüye ait anlamsal bilgiler (sınıflandırma, bölütleme) içermektedir. Anlamsal boşluk sorunu, sığ katmanlardan elde edilen özellikler ile derin katmanlardan elde edilen özellikler arasında oluşmaktadır.

Bu problemin çözümü için önerilen bir yöntemde (Pang, vd., 2019) farklı seviyelerdeki özellikler Paralel Piramit (Parallel Pyramid) yöntemi (farklı seviyelerdeki özelliklerin birleştirilmesi) ve aşağıdan-yukarıya (bottom-up) yaklaşımıyla (sığ katmanlardaki özelliklerden derin katmanlardaki özelliklere doğru ilerleme) anlamsal bölütleme işlemi için kullanılmıştır.

Görüntü içeriği ve görüntüyü tanımlayan sınıf etiketi arasında var olan anlamsal boşluk sorununun çözümünde geliştirilen çerçevede (Ma, vd., 2010), görüntülerden çıkarılan özellikler ile Görüntü Benzerlik Çizgesi (Image Similarity Graph) oluşturulmuş ve geliştirilen Rastgele Yol Modeli (Random Walk Model) isimli algoritma sayesinde iki düğüm arasındaki yol uzunluğu bilgisi hesaplanmıştır. Böylece, görüntü içeriği ile görüntüyü tanımlamak için kullanılan sınıf etiketleri arasındaki anlamsal boşluğun kapatılması için etkili sonuçlar elde edilmiştir.

Tıbbi İçerik Tabanlı Görüntü Erişimi alanında karşılaşılan sistem performansını ve kullanılabilirliğini etkileyen Anlamsal Boşluk kavramına ait yapılan bir sistematik analiz ve sınıflandırma çalışmasında (Deserno, vd., 2009), sorunun çözümünde kullanılan İçerik (Content: görüntülerin insanlar tarafından algılanan biçimi), Özellikler (Features: görüntülerden çıkarılan düşük düzeydeki renk, şekil, doku özellikleri), Performans (Performance: sistem entegrasyonu ve değerlendirilmesi), Kullanılabilirlik (Usability: sistemin kullanılabilirliği) kategorileri üzerinden bir tablo yardımıyla karşılaştırma yapılmıştır.

Anlamsal Boşluk başlığı altında incelenen makalelerde kullanılan Yöntem, Veri kümesi, Başarım ölçeği ve Sonuçlara Tablo 4'te yer verilmiştir.

Tablo 4

Anlamsal Boşluk başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

Makale	Yöntem	Veri kümesi	Başarım ölçeği	Sonuç
(Zhao and Grosky 2002)	Gizli Anlamsal İndeksleme (Latent Semantic Idexing - LSI)	Web	Precision (Hassasiyet), Recall (Geri çağırma)	0.35 0.88
(Pang, vd., 2019)	Paralel Özellik Piramidi (Prallel Feature Pyramid - SeENet)	Pascal VOC 2012 CamVid Cityscapes	Ortalama Birleşim Üzerinde Kesişim (Mean pixel-intersection over union-mIoU)	86.6 68.4 81.2
(Ma, vd., 2010)	Rastgele Gezinti Modeli (Random Walk Model)	Flickr	N-Hassasiyet (Precision@N)	0.68

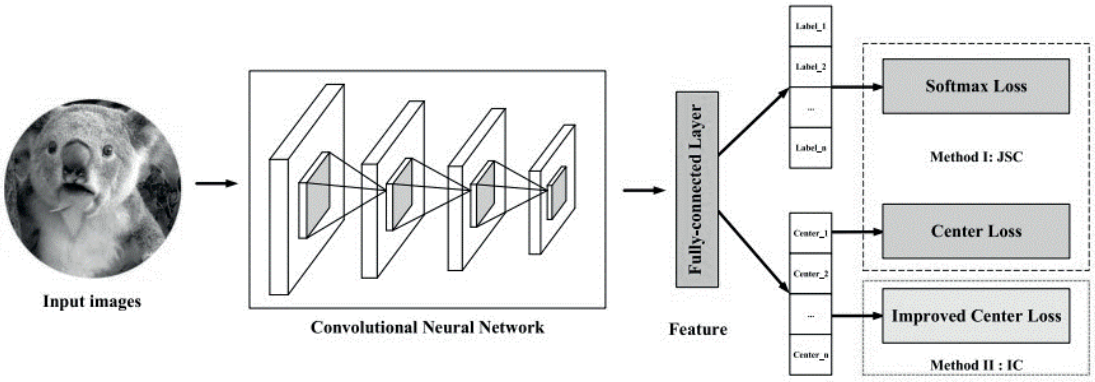
4. ANLAMSAL TABANLI GÖRÜNTÜ ERİŞİMİ

Anlamsal Tabanlı Görüntü Erişimi'nin amacı, anlamsal benzer içeriğe sahip resimlerin bir veritabanında aranmasıdır. SBIR yaklaşımında image annotation işlemiyle, görüntüleri tanımlayan bir sınıf etiketinin eklenmesi yapılmaktadır. Sonrasında, image retrieval işleminde verilen etiket ile eşleşen görüntülerin aranması gerçekleştirilmektedir (Carneiro, vd., 2007)

SBIR ile anlamsal boşluk olarak adlandırılan ve resimlerden çıkarılan düşük düzeydeki özellikler ile insanlar tarafından görüntüden algılanan yüksek düzeydeki kavramlar arasındaki açığın kapatılması yönünde çalışmalar yürütülmektedir (Liu, vd., 2007). Bu alandaki bir çalışmada (H. H. Wang, Mohamad, ve Ismail 2010), Anlamsal Tabanlı Görüntü Erişiminde kullanılan teknikler, bu alanda yapılan araştırmalar ve bu araştırmalarda kullanılan yaklaşımlar bir tablo şeklinde özetlenmiş ve Anlamsal Tabanlı Görüntü Erişimi yaklaşımı için genel bir çerçeve sunulmuştur. Ayrıca, Görüntü Erişimi yaklaşımında özellik çıkarımının öneminden bahsedilmiş ve araştırmanın sonucunda çalışmaların düşük düzeydeki özelliklerin çıkarılmasından, anlamsal özelliklerin çıkarılmasına doğru yöneldiği belirtilmiştir.

Yapılan diğer bir çalışmada (Su, vd., 2011), görüntülerden çıkarılan özellikler ve insanların bu görüntülerden algıladığı kavramlar arasındaki ilişkiler temel alınarak düzensizlik (entropy), terim frekansı-belge frekans tersi (term frequency-inverse document frequency) ve ilişkilendirme kuralları (association rules) yöntemlerinin entegrasyonu ile SVM ve k-Means Clustering kullanlarak image annotation işlemi yapılmıştır.

Özellik çıkarımı, SBIR'de önemli bir rol oynamaktadır. Özellik çıkarımı için en sık kullanılan mimari Evrişimli Sinir Ağları Mimarisi'dir. Önerilen çalışmada (Song, vd., 2018), Evrişimli Sinir Ağları kullanılarak görüntülerden özellikler çıkarılmış, gelişitrilen kayıp fonksiyonu (loss function) ile Evrişimli Sinir Ağı eğitilerek parametrelerin optimizasyonu yapılmış ve snıflandırma işleminin doğruluk oranında (classification accuracy) iyileştirme elde edilmiştir. Sistemin mimarisi, Şekil 7' de gösterilmiştir.



Şekil 7. Önerilen çalışmada (Song, vd., 2018) kullanılan Ayrık Anlamsal Derin Öğrenme (Discriminative Semantic-based Deep feature Learning) Mimarisi

Farklı derin öğrenme mimarilerinin bir arada kullanılması ile insanlar tarafından görüntülerden algılanan yüksek düzeydeki anlamsal özelliklerin çıkarılması yönünde çalışmalar da yapılmaktadır. Yüksek düzeydeki özelliklerin çıkarılması amaçlı yapılan bir çalışmada (Wu, 2020), iki farkli derin öğrenme mimarisi Derin Boltzman Makineleri (Deep Boltzman Machine-DBM) ve Evrişimli Sinir Ağları birlikte kullanılmıştır. CNN ile düşük düzeydeki özelliklerin çıkarımı yapılmış, DBM ile görüntüye ait yüksek düzeydeki ilişkiler oluşturulmuştur. Özellik çıkarımından sonraki adımda, sınıflandırma amaçlı kullanılan Seyreltilmiş Softmax (Dropout Regularized Softmax) fonksiyonu ile STL-10 veri kümesi üzerinde %60,3 başarım elde edilmiştir.

Anlamsal Tabanlı Görüntü Erişimi işleminin başarımını arttırmak için "Ontolojiler" kullanılarak yüksek düzeydeki kavramların tanımlanması yapılmaktadır. Varlıkların tanımlanması için sınıflar oluşturulur ve sınıflar arasında ilişkiler belirlenir (Parsons, 2009). Görüntü erişimi işlemlerinde, bazen iki resim özellikleri arasında benzerlik olsa da anlamsal farklar bulunmaktadır. Özellikle, Yazı Tabanlı Görüntü Erişiminde sorguda kullanılan ifade ile sonuç olarak getirilen görüntüler arasında uyuşmazlıklar olmaktadır.

Önerilen çalışmada (Bouchakwa, vd., 2020), Çok Seviyeli Görüntü Çeşitlendirme Süreci (Multi-level Image Diversification Process) olarak ifade edilen teknikle, anlamsal benzer resimler aynı küme içinde toplanmış ve sorgu ifadesine göre filtrelenerek yeniden sıralanmıştır. Sorguda kullanılan arama ifadelerinin farklı anlamları olabilmektedir. Bu problem, Belirsiz Sorgu Sorunu (Query Ambiguous Problem - QAP) olarak ifade edilir. Aynı çalışmada, QAP probleminin çözümü için sorguda kullanılan ifadelerin farklı anlamları, oluşturulan Ontolojilere göre aranmış ve SPARQL (Anlamsal web sorgulama dili) ile sorgu ifadesine benzer resimlerin aranması işlemi gerçekleştirilmiştir.

TBIR işleminde resimlere ait içerik, resim yazılarıyla (caption) ifade edilmektedir. Resimleri tanımlayan yazıların Kaynak Betimleme Çerçevesi (Resource Description Framework - RDF) ile gösterildiği bir çalışmada (Chen, vd., 2020), resim içeriğine ait nesneler, özellikleri ve ilişkileri önce Doğal Dil İşleme (Natural Language Processing - NLP) ile çıkarılmış ve Kaynak Betimleme Çerçevesi kullanılarak oluşturulan yönlü çizge (directed graph) ile gösterilmiştir. Sonrasında SPRAQL sorgulama dili kullanılarak arama işlemi gerçekleştirilmiştir. Ontolojilerin kullanıldığı diğer bir çalışmada (Mezaris, vd., 2003), önce resimler bölgelere ayrılmış, resimlerden çıkarılan özellikler oluşturulan ontolojiler ile bağdaştırılmıştır. Sonra, yüksek düzeydeki kavramların tanımlanması amacıyla oluşturulan ontolojilere göre anahtar kelimeler belirlenmiştir. Anlamsal Tabanlı Görüntü Erişimi uygulamalarında, arama işleminin iyileştirilmesi amacıyla Anlamsal Ağ'ın (Semantic Web) getirdiği yenilikler (Ontology, RDF) kullanılmaktadır.

Anlamsal Ağ ile günümüzde kullanılan web'in temel yaklaşımı olan Belgelerin Ağı (Web of Document), belgelerin bağlantılar yardımıyla birbirine bağlanması yaklaşımından Bilgilerin Ağı (Web of Data), web üzerinde aralarında ilişkiler tanımlanmış bilgiler yaklaşımına doğru bir geçiş söz konusudur. Anlamsal Ağda bilgiler, bilgisayarların işleyebileceği forma dönüştürülmekte

ve bilgiler arasında anlamsal ilişkiler oluşturulabilmektedir. Yapılan bir çalışmada (Li, vd., 2007) önerilen "OntoLook" isimli sistem ile web sayfalarının içeriği, anlamsal ağda kullanılan ifadeler ile (Web Ontology Language - OWL, Resource Description Framework - RDF) yapılandırılmış ve anahtar kelime temelli bir arama motoru oluşturulmuştur. Yapılan diğer bir çalışmada (Minu ve Thyagharajan, 2014) ise "çiçek bilgi alanı – asteroideae flower" üzerinde Çok Seviyeli Görüntü Ontolojisi (Multi-level image ontology) yaklaşımı ile yazı ve görsel özelliklerin kombinasyonuyla, düşük düzeydeki özelliklerin anlamlandırılması yapılmıştır.

Anlamsal Tabanlı Görüntü Erişimi başlığı altında incelenen makalelerde kullanılan Yöntem, Veri kümesi, Algoritma, Başarım ölçeği ve Sonuçlara Tablo 5'te yer verilmiştir.

Tablo 5

Anlamsal Tabanlı Görüntü Erişimi başlığı altında incelenen çalışmalar ve elde edilen sonuçlar

Makale	Yöntem	Veri kümesi	Başarım ölçeği	Sonuç
(Song, vd., 2018)	Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)	ILSVRC Caltech256	Ortalama hasssasiyet (Mean Average Precision - MAP), Geri çağırma (Recall)	0.608 0.789 0.502 0.507
(Wu 2020)	Evrişimli Derin Boltzman Makineleri (Convolutional - Deep Boltzman Machine C-DBM)	STL-10	Hassasiyet (Precision) Geri çağırma (Recall) F değeri (F value)	63.69 62.62 0.6315
(Bouchakwa, Ayadi, ve Amous 2020)	Yazı Tabanlı Görüntü Alma (Text Based Image Retrieval -TBIR)	Flickr	Hassasiyet (Precision) Geri çağırma (Recall) F değeri (F value)	86.83 83.10 84.93
(Chen, vd., 2020)	Doğal Dil İşleme Natural Language Processing - NLP), Kaynak Betimleme Çerçevesi (Resource Description Framework - RDF	Flıckr8k	F değeri (F-score)	0.9091
(Mezaris, Kompatsiaris, ve Strintzis 2003)	Resim Bölütleme Algoritması (Image Segmentetion Algorithm)	Corel	Hassasiyet (Precision) Geri çağırma (Recall)	0.60 0.80
(Li, Wang, ve Huang 2007)	Yazı Tabanlı Görüntü Alma (Text Based Image Retrieval -TBIR)	Web	Zaman karmaşıklığı (Time complexity) Eşik (Threshold) yok - CPU ms Eşik (Threshold) var - CPU ms	O(2n) 3125 500
(Minu ve Thyagharajan 2014)	K-Ortalama kümeleme (K-means clustering)	Hortipedia's flower dataset	Üstün Standart Ölçüm (Golden standard measure)	%72

5. SONUÇ

Anlamsal Tabanlı Görüntü Erişimi işleminin amacı, görüntülerden çıkarılan düşük düzeydeki özellikler renk, şekil, doku ile insanlar tarafından görüntüden algılanan yüksek düzeydeki kavramlar arasında oluşan ve Anlamsal Boşluk olarak ifade edilen darboğazın aşılmasıdır. Kısaca, SBIR anlamsal benzerliğe sahip olan resimlerin aranması ile ilgilidir. Görüntüye ait özelliklerin çıkarılmasında ve bölütleme (segmentation) işleminde, derin öğrenme mimarileri kullanılmaktadır. CNN mimarileri ile resimlerden çıkarılan sığ katmanlardaki özellikler görüntüye ait detaylı bilgileri içerirken, derin katmanlardaki özellikler ise görüntüye ait anlamsal bilgiler içermektedir.

Anlamsal Tabanlı Görüntü Erişimi ile anlamsal boşluk probleminin çözümünde kullanılan yöntemlerin başında, Ontoloji kavramının kullanılması gelmektedir. Ontolojiler ile insanlar tarafından görüntülerden algılanan yüksek düzeydeki kavramların tanımlanması yapılmaktadır. Anlamsal ağ kavramının gelişmesiyle, Ontolojiler kullanılarak bilgiler bilgisayarların işleyebileceği biçime dönüştürülmekte ve bilgiler arasında anlamlı ilişkiler oluşturulabilmektedir. Ontolojiler ile tanımlanmış kavramlardan yararlanılarak, Anlamsal Tabanlı Görüntü Erişimi işleminin anlamsal boşluk sorununun çözümünde etkili sonuçlar verdiği görülmektedir.

Görüntüler içerisinden nesnelerin algılanmasından ve bölütleme işleminin yapılmasından sonra, algılanan nesneyi tanımlayacak bir sınıf etiketinin atanması gerekmektedir. Anlamsal Boşluk sorununun çözümü, nesnelerin tanımlanması işlemi için belirli

bir bilgi alanı (domain) üzerinde oluşturulacak olan Ontolojiler ile gerçekleştirilebilir. Nesnelerin tanımlanması amacıyla, eklenecek kelimelerin farklı anlamlarının da kapsanması için geliştirilen WordNet isimli sözlük veritabanı olarak kullanılabilir.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Author Contributions: Conception/Design of Study- A.G.,M.D.,T.E.; Data Acquisition- A.G.; Data Analysis/Interpretation- M.D.,T.E.; Drafting Manuscript- A.G.,M.D.,T.E.; Critical Revision of Manuscript- M.D.,T.E.;Final Approval and Accountability- A.G.,M.D.,T.E.
Grant Support: The authors declared that this study has received no financial support.

Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.
Yazar Katkıları: Çalışma Konsepti/Tasarım-A.G.,M.D.,T.E.; Veri Toplama- A.G.; Veri Analizi/Yorumlama- M.D.,T.E.; Yazı Taslağı- A.G.,M.D.,T.E.; İçeriğin Eleştirel İncelemesi- M.D.,T.E.; Son Onay ve Sorumluluk- A.G.,M.D.,T.E.
Finansal Destek: Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

Kaynaklar/References

Alkhawlani, 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).

Alpaydın, E. (2004). *Introduction to machine learning*. MIT Press.

Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, 32, 20–54. https://doi.org/10.1016/j.jvcir.2015.07.012

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

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

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

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.2016.2644615

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

Chen, H., Guo, A. Bin, Ni, W., & Cheng, Y. (2020). Improving the representation of image descriptions for semantic image retrieval with RDF. *Journal of Visual Communication and Image Representation*, 73(August 2019), 102934. https://doi.org/10.1016/j.jvcir.2020.102934

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

Deserno, T. M., Antani, S., & Long, R. (2009). Ontology of gaps in content-based image retrieval. *In Journal of Digital Imaging* (Vol. 22, Issue 2, pp. 202–215). https://doi.org/10.1007/s10278-007-9092-x

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *In Neurocomputing* (Vol. 187, pp. 27–48). https://doi.org/10.1016/j.neucom.2015.09.116

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

Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1), 262–282. https://doi.org/10.1016/j.patcog.2006.04.045

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *In IEEE*. https://doi.org/10.1109/CVPR.2015.7298965

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

Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2003). An ontology approach to object-based image retrieval. *Proceedings 2003 International Conference on Image Processing* (Cat. No.03CH37429), 3, II-511–514. https://doi.org/10.1109/ICIP.2003.1246729

Minu, R. I., & Thyagharajan, 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

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

Noh, H., Hong, S., & Han, B. (2015). Learning Deconvolution Network for Semantic Segmentation (Vol. 1). https://doi.org/10.1109/ICCV.2015.178

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

Parsons, S. (2009). A Semantic Web Primer, Second Edition by Antoniou Grigoris and Harmelen Frank van, MIT Press, 288 pp.. *In The Knowledge Engineering Review* (Vol. 24, Issue 4). https://doi.org/10.1017/s0269888909990117

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

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380. https://doi.org/10.1109/34.895972

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

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

Wang, Q., Lai, J., Claesen, L., Yang, Z., Lei, L., & Liu, W. (2020). A novel feature representation: Aggregating convolution kernels for image retrieval. Neural Networks, 130, 1–10. https://doi.org/10.1016/j.neunet.2020.06.010

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

Zhang, Y., Sidibé, D., Morel, O., & Mériaudeau, F. (2020). Deep multimodal fusion for semantic image segmentation: A survey. *Image and Vision* Computing, 104042. https://doi.org/https://doi.org/10.1016/j.imavis.2020.104042

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

Zhu, H. (2020). Massive-scale image retrieval based on deep visual feature representation. *Journal of Visual Communication and Image Representation*, 70. https://doi.org/10.1016/j.jvcir.2019.102738

WordNet. Retrieved from https://wordnet.princeton.edu, (22.11.2020)