

# GÖRSEL ARAMA SONUÇLARININ ÇOKLU ÖRNEKLE ÖĞRENME YÖNTEMİYLE YENİDEN SIRALANMASI

## MULTIPLE INSTANCE LEARNING FOR RE-RANKING OF WEB IMAGE SEARCH RESULTS

Fadime Şener<sup>1</sup>, Nazlı İkizler Cinbiş<sup>2</sup>, Pınar Duygulu Şahin<sup>1</sup>

<sup>1</sup> Bilgisayar Mühendisliği Bölümü  
Bilkent Üniversitesi

[fadime.sener@cs.bilkent.edu.tr](mailto:fadime.sener@cs.bilkent.edu.tr), [duygulu@cs.bilkent.edu.tr](mailto:duygulu@cs.bilkent.edu.tr)

<sup>2</sup> Bilgisayar Mühendisliği Bölümü  
Hacettepe Üniversitesi

[nazli@cs.hacettepe.edu.tr](mailto:nazli@cs.hacettepe.edu.tr)

### ÖZETÇE

*Bu çalışmada, çoklu öğrenme yöntemi ile metin tabanlı arama motorlarından elde edilen görsel sorgu sonuçlarını iyileştirmek için geliştirilmiş olan, zayıf denetimli öğrenen bir yöntem sunulmaktadır. Bu yöntemde arama motorundan dönen sonuçlar zayıf pozitif kabul edilerek, sorgu kategorisinden görüntü içermeyen negatif görüntüler de kullanılarak; çoklu örnekle öğrenme için torbalar oluşturulmaktadır. Bu torbalar ve veri kümesindeki örnekler arasında kurulan torba-örnek benzerliğinden yararlanarak; torbalar yeni bir örnek uzayına taşınmakta ve problem klasik bir denetimli öğrenme problemi haline getirilmektedir. Daha sonra, lineer destek vektör makinesi (DVM) kullanılarak her sorgu için sınıflandırma modelleri oluşturulmaktadır. Elde edilen sınıflandırma değerlerine göre görseller yeniden sıralanmış ve arama motorundan gelen sonuçların iyileştirildiği görülmüştür. Bu çerçevede, torba boyları arasında bir örüntü bulmak için yaptığımız deneyleri sunmaktayız.*

### ABSTRACT

*In this study, we propose a weakly-supervised multiple instance learning (MIL) method to improve the results of text-based image search engines. In this approach, ranked image list of search engine for a keyword query is treated as weak-positive input data, and with additional negative input data, multiple instance learning bags are constructed. Then, Multiple Instance problem is converted to a standard supervised learning problem by mapping each bag into a feature space defined by instances in training bags using a bag-instance similarity measure. At the end, linear SVM is used to construct a classifier to re-rank keyword-based image search data. Based on the classification scores, we re-rank the images and improve precision over the search engine results. In this respect, we also present our experiments conducted to find a pattern for multiple instance bag sizes to obtain better average precision.*

### 1. GİRİŞ

Günümüzde internet gelişimini hızla sürdürürken, barındırdığı veri miktarı da hızla artmaktadır. Kullanıcılar, bu verilere erişmek için Google, Bing, Yahoo vb. arama

motorlarını yaygın bir şekilde kullanmaktadır. Bu süreçte arama motorlarının metin tabanlı aramalarda, ilgili verileri döndürme oranı da yapılan birçok çalışma ile beraber yüksek başarılar göstermektedir; fakat görsel aramalarda ilgili sonuç verme başarısı metin başarısına göre nispeten daha düşüktür. Bunun sebebi ise arama motorlarının görsel sorgulara, görüntü tabanlı değil de, metin tabanlı olarak yaklaşmasıdır. Görsel arama sonuçlarının başarısının düşük olmasının sonucu olarak örneğin; Google arama motorundaki görsel arama sonucunda, kullanıcılar ilk sayfalarda sorgu ile ilgisiz görüntülerle karşılaşabilmektedirler. Bu durum daha iyi bir sıralama ihtiyacını doğurmaktadır.

Metin tabanlı görüntü arama sonuçlarını iyileştirmek için; nesne, sahne tanıma yöntemlerini kullanan çalışmalar mevcut olmasına rağmen, hala bir görselin aranan görüntü olup olmadığına karar verme, oldukça zor bir bilgisayarlı görme problemidir, çünkü görüntüler aydınlanma, arka plan dokusu, bakış açısı değişikliği vb. açısından oldukça farklı varyasyonlara sahip olabilir.

Bu çalışmadaki amacımız metin tabanlı arama sonuçlarının, görüntünün içeriğine dayalı olarak yeniden sıralanması ile ilgisiz görüntülerin ilk sayfalarda görünmesini engellemek, resim erişim başarımını yükseltmektir. Bizim bu çalışmadaki temel katkımız ise; bu probleme bir çoklu örnekle öğrenme problemi gibi yaklaşmamız ve bu yöntemimizin bir sonucu olarak, mevcut çalışmalardan farklı olarak etiketli veriye ihtiyaç duyulmamasıdır. Etiketleme, insan çabası gerektiren zor bir iş olduğu için çoğu araştırmacının az sayıda veri kümesine bağımlı kalmasına neden olur. Öte yandan, etiketleme gerekmeyen durumlarda, çalışma çok daha yüksek boyuttaki veriye ölçeklenebilir olmaktadır.

Önerilen yöntem, zayıf-denetimli öğrenme tabanlı olarak pozitif ve negatif örnek torbaları oluşturmaktadır. Pozitif örnek torbaları, bir sorgu için metin tabanlı arama motorundan dönen görüntülerin sırası göz önüne alınarak, herhangi bir etiketleme gerektirmeden oluşturulmaktadır. Öğrenme için kullanılacak negatif görüntüler ise; o sorgu ile ilgisi olmayan örneklerden elde edilmektedir. Çoklu örnekle öğrenme kullanılarak, görüntülerin ilgili veya ilgisiz olup olmadıkları belirlenmekte ve oluşturulan sınıflandırma sonuçlarına göre arama motoru sonuçları yeniden sıralanmaktadır. Böylece önerilen yöntem, kullanıcıya, ilgili görüntüleri ilk sayfalarda sunmaktadır

## 2. İLGİLİ ÇALIŞMALAR

Çoklu örnekle öğrenme geleneksel öğrenme yöntemlerinden farklı olarak, etiketlerin tek tek örnekler yerine örnek torbalarına atandığı gözetimli bir öğrenme yöntemidir. Başka bir deyişle, bu öğrenme yönteminin verileri örnek torbalarıdır. Örnek torbalarının etiketleri bilinirken, içerdikleri örneklerin teksel etiketleri hakkında bilgi yoktur. Bununla beraber torba oluşturulurken iki varsayım mevcuttur; bir torbanın pozitif olarak etiketlenmesi için en az bir pozitif örnek içermesi gerekirken, negatif olarak etiketlenmesi için içerdiği tüm örneklerin negatif olması gerekir.

Çoklu örnekle öğrenme, son yıllarda veri madenciliği, sahne öğrenme, metin sınıflandırma gibi birçok uygulamada kullanılmaktadır. Bu yöntemin ilk adımı 1990 yılında Keeler ve arkadaşları el yazısı ile yazılmış sayıların en iyi bölütlemesini bulabilmek için bir nöron ağı tasarladıklarında atılır[1]. Bu çalışmada etiket sayı görüntüsünün tamamına değil de, o görüntünün her bölümünün toplamına verilmiştir. Bu fikir daha sonra Dietterich ve arkadaşları tarafından çoklu örnekle öğrenme olarak isimlendirilmiş ve bir ilacın etkisini tahmin etme problemi için kullanılmıştır [2]. Bu çalışmada her molekül bir örnek torbası olarak kabul edilip, o molekülün varyantları ise torbadaki örnekler olarak kullanılarak, hedef proteinle bağlanıp bağlanamaması belirlenmeye çalışılmıştır.

Bir başka çoklu örnekle öğrenme uygulaması ise Değişik Yoğunluk (Diverse Density) çatısıdır[3]. Bu uygulamada nitelik uzayındaki örnekler için Diverse Density(DD) fonksiyonu tanımlanır. Bu fonksiyon, farklı örnek torbalarındaki örneklerin, aynı etiket ile bulunup bulunmamasını ölçer. Belli bir eşik değerinden fazla DD'ye sahip olan noktaların tüm pozitif torbaların en az bir noktasına yakın olduğunun ve tüm negatif torbalardan çok uzak olduğunun göstergesidir. Eğer en yüksek DD'ye sahip bir nokta bulunursa, yeni test örnekleri artık bu noktaya olan uzaklıklarına göre sınıflandırılabilir.

Bu makalede kullandığımız yöntem ise Chen ve arkadaşları [4] tarafından DD çatısına dayalı olarak geliştirilen yöntemdir. Kısa adı MILES olan bu yöntem, çoklu örnekler içinde gömülü örnek seçimine bağlı olarak yapılır.

## 3. YÖNTEM

Yöntemimizde, metin tabanlı arama motorlarından metin sorgusu sonucunda dönen resimleri otomatik olarak yeniden sınıflandırmaya çalışmaktayız. Bu amaçla, sorgu sonucu dönen resimlerden, öncelikle görsel öznitelikler çıkararak veri kümesi oluşturulur, ardından her veri örneği, (instance) gelen resim sırası göz önünde bulundurularak veri torbalarına otomatik olarak atanır. Daha sonra bu veri torbaları üzerinde çoklu örnekle öğrenme yöntemi ile sınıflandırıcılar oluşturulur ve bu sınıflandırıcı sonuçlarına göre görseller yeniden sıralanır. Yöntemin detaylı aşamaları aşağıda sunulmaktadır.

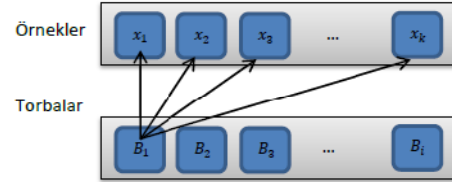
### 3.1. Nitelik Çıkarma ve Öğrenme

Bu çalışmada her görüntü, görsel kelime kümeleri (bag of visual words) [5] ile ifade edilmektedir. Bu amaçla, öncelikle her resimden  $x, y$  koordinatlarını, ölçeği ve kontrast değerlerini içeren  $[4xN]$  boyutunda ( $N$  tane ilgi noktası bulunmaktadır) SIFT ilgi noktaları matrisi çıkarılmaktadır [6]. Ardından bu  $N$  tane ilgi noktasının  $[128x1]$  boyutlu gradyen öznitelikleri hesaplanır ve elde edilen  $[128xN]$  boyutunda öznitelik matrisi,  $k$ -means kümeleme yöntemi kullanılarak 1000 gruba kümelendir. Bu kümelerin orta noktaları görsel kelimeleri

oluşturmaktadır. Bundan sonra görüntü için hesaplanacak kelimelerin, sadece görüntü içerisinde var olup olmadıkları değil, görüntünün neresinde bulduklarının bilgisini de almak için, görüntü  $4x4$  'lük uzamsal parçalara bölünür. Ardından her görüntü için 16 farklı uzamsal parçadan çıkarılan SIFT öznitelik vektörlerinin, hesaplanmış kelime noktalarına Öklid uzaklığı hesaplanır ve her öznitelik vektörü, kendisine en küçük uzaklığa sahip görsel kelime ile eşleştirilir. Bu görsel kelime kümelerinin görülme sıklığı hesaplanarak kelime kümesi dağılımı (histogramı) hesaplanır ve her uzamsal parça  $[1x1000]$  boyutlu dağılımlar ile ifade edilir. Bu aşamaların sonucunda her bir resim için 16 farklı uzamsal parçanın dağılımlarının birleşiminden oluşan  $[1x16000]$  boyutlu nitelik vektörü elde edilir.

### 3.2. Çoklu Örnekle Öğrenme

Görsellerin sırasını zayıf eğitici olarak, etiketlemeye gerek kalmadan öğrenmek için MILES [4] yöntemini kullanabiliriz. Bu yöntemde, pozitif ve negatif torbalar, torbaları oluşturmak için kullanılan tüm örneklerin oluşturduğu yeni bir uzayda temsil edilirler. Bunun için her torba bir benzerlik fonksiyonu kullanılarak yeni uzaya taşınır. Bir torba ve bir örnek arasındaki benzerlik, torbadaki tüm örneklerden bu örnek ile maksimum benzerliğe sahip olanın benzerliği seçilerek belirlenir. Şekil 1 torbalar ile tüm örnekler arasında kurulan benzerliği göstermektedir.



Şekil 1:  $x_i$  torbaları oluşturan örnekler,  $B_j$  örneklerden oluşan torbalar.

Formal olarak kullanılan notasyon şu şekildedir: Pozitif torbalar  $B_i^+$  ile ve bu torbadaki  $j$ 'inci örnek  $x_{ij}^+$  ile gösterilir.  $B_i^+$  torbası  $n_i^+$  tane örnek içerir;  $x_{ij}^+$ ,  $j = 1, 2, \dots, n_i^+$ . Benzer şekilde negatif torbalar da  $B_i^-$ ,  $x_{ij}^-$  ve  $n_i^-$  ile simgelenir.  $B_i$  ise torbanın negatif veya pozitif sunumundan ve örneklerden bağımsız, torba gösterimidir. Böylece bir torba  $B_i$  ve bir örnek  $x_k$  arasında benzerlik aşağıdaki formül ile ifade edilir.

$$s(x_k, B_i) = \max_j \exp\left(-\frac{\|x_{ij} - x_k\|}{\sigma}\right) \quad (1)$$

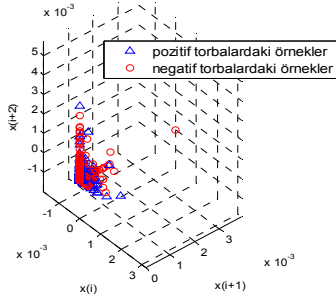
Bu benzerliği kullanarak, yeni uzaydaki torba sayısı  $l$  ise, (hem  $l^+$  hem de  $l^-$  torbaları içerir) yeni uzayda bütün torbalar;

$$\begin{bmatrix} s(x_1, B_1) & \dots & s(x_1, B_l) \\ s(x_2, B_1) & \dots & s(x_2, B_l) \\ \dots & \dots & \dots \\ s(x_k, B_1) & \dots & s(x_k, B_l) \end{bmatrix}$$

olarak ifade edilir. Burada her kolon bir torbayı ifade ederken, satırlar da elimizdeki  $k$  tane örneği göstermektedir.

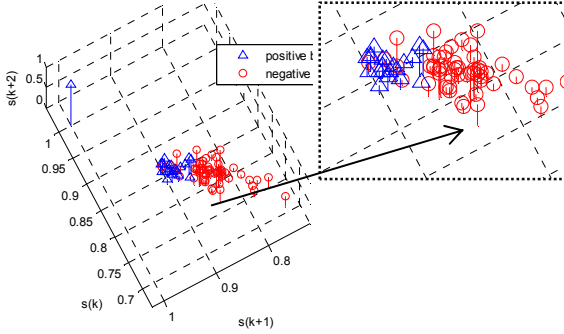
**Örnek:** Elimizdeki, arama motorundan elde edilen ve pozitif kabul ettiğimiz 651 örnek ve 1659 negatif örnek içeren bir veri kümesinden, rastgele boylarda torbalar oluşturduk. Şekil 2 tüm örneklerin üç boyutlu nitelik uzayındaki gösterimidir. Her görüntü  $[1x16000]$  nitelik vektörü ile tanımlıdır. Dikkat edilirse pozitif ve negatif torbalar bu düzlemde birbirlerine çok yakın görülmektedir, çünkü zayıf-öğreticili bu yöntemde, etiketler kesin değildir ve pozitif torbalar negatif örnekler içerebilir.

Eğitim kümesi: 33 pozitif torba + 83 negatif torba



Şekil 2: Veri kümesindeki örneklerin iki boyutlu nitelik uzayında görünümü

Şekil 3'te bu torbaların, veri kümesindeki bütün örnekler ile benzerliklerinden elde edilen, yeni uzaydaki 3 boyutlu görünümü sunulmuştur. (3 örnek boyut rastgele seçilmiştir.) Dikkat edilirse, uyguladığımız bu yöntemle pozitif ve negatif torbalar bir düzlemle ayrılabilir duruma gelmiştir. Torbalar ayrılabilir olmakla beraber pozitif etiketlenmiş torbaların yakınında negatif torbalar da oluşmaktadır, bu da yöntemimizin zayıf-öğreticili olmasının bir sonucudur, arama motorundan dönen tüm örnekleri pozitif kabul ettiğimiz için pozitif etiketli negatif torbalar oluşmuştur.



Şekil 3: Torbaların yeni uzayda 3 boyutlu görünümü.

### 3.3. Torba Oluşturma

Arama motorundan metin tabanlı sorgularla dönen resimlerin bir bölümü, aranan sorguya ait olan (pozitif), bir bölümü ise sorguya ait olmayan (negatif) resimlerdir. Çoklu örnekle öğrenme algoritmasını bu resimlere uygulamak için, örnekler içinden otomatik olarak pozitif torbaları oluşturmak gerekmektedir. Bu amaçla, sorgu için gelmiş olan resimler arasından  $k$  tane örnek içeren torbalar oluşturulur. Öte yandan, negatif torbaları oluşturmak için de bu sorgu dışındaki sorguların sonuçları kullanılır.

Torbaları oluşturmada, torba içindeki örnek sayısı ( $k$ ) önemli bir parametredir. Örneklerin esas etiketleri bilinmediği için, farklı değerlerdeki örnek sayısı kullanmak, başarıyı etkileyebilir. Zayıf denetimli bu öğrenme yönteminde,  $k$  değeri seçilirken şu soruya cevap aranmaktadır: Oluşturulan pozitif torba boyu ne olmalı ki, her pozitif torbaya en az bir pozitif örnek düştüğünden emin olalım? Bu nedenle torbaları rastgele oluşturmak yerine, resimlerin arama motorunun getirdiği sırada kullanılması daha avantajlı olabilir. Bu amaçla, bu çalışmadaki deneylerde, farklı  $k$  değerleri ile resimlerin arama motorundan dönüş sırası esas alınarak oluşturulan torbalar üzerinde, çoklu örnekle öğrenmenin, yeniden sıralamanın başarımını nasıl etkilediği incelendi.

Torbalar oluşturulduktan sonra, lineer destek vektör makineleri (DVM) kullanılarak sınıflandırma modeli

hesaplanmaktadır. Test aşamasında arama motorundan dönen her bir görüntü, tek başına bir torba olarak varsayılmakta ve oluşturulan sınıflandırıcılar her torbaya uygulanmaktadır. Elde edilen sınıflandırma değerlerine göre görseller yeniden sıralanmaktadır.

## 4. VERİ KÜMESİ

Bu çalışmada metin tabanlı Google arama motorunun görseller arasındaki arama sonuçlarını içeren Schroff vd.'nin [6] veri kümesini kullanıldı. Yöntem 7 farklı sorgu için test edildi, bunlar "airplane", "bikes", "kangaroo", "penguin", "shark", "tiger" ve "zebra" sorgularıdır. "Airplane" sorgusu için veri kümesinde 275 pozitif, 376 negatif, "bikes" sorgusu için 293 pozitif, 297 negatif, "kangaroo" için 214 pozitif, 437 negatif, "penguin" için 326 pozitif, 375 negatif, "shark" için 221 pozitif, 491 negatif, "tiger" için 163 pozitif, 480 negatif, "zebra" için ise 254 pozitif, 465 negatif resim bulunmaktadır

## 5. DENEYLER

Sonuçlarımızı değerlendirirken kesinlik (precision) ve geri getirme (recall) oranlarını kullandık. Kesinlik, kullanılan yöntemin tahmin ettiği doğru pozitiflerin oranı iken, geri getirme pozitif görüntülerin döndürülme oranıdır. Tablo 1'de farklı örnek sayısına bağlı olarak torba oluşturma sonuçları verilmiştir. Bu tabloda,  $k_p$  pozitif torbalardaki örnek sayısını,  $k_n$  ise negatif torbalardaki örnek sayısını göstermektedir.

Tablo 1: Negatif torba boyu sabit  $k_n = 10$  iken,  $k_p=5,10,15,20,30$  boyları ile oluşturulan pozitif torbaların, Google sorgusu sonuçları ile precision-recall eğrisi altında kalan alan bakımından karşılaştırılması (yüzde olarak %).

Sorgu	Google	$k_p=5$	$k_p=10$	$k_p=15$	$k_p=20$	$k_p=30$
airplane	50.91	52.65	<b>53.92</b>	50.51	47.57	46.20
bikes	51.60	59.95	<b>62.95</b>	56.02	52.56	51.20
kangaroo	43.67	49.13	<b>52.22</b>	43.04	30.22	28.03
penguin	53.89	<b>54.84</b>	54.54	50.07	49.23	49.07
shark	35.71	25.71	46.19	47.41	<b>48.61</b>	47.00
tiger	34.51	40.67	<b>49.07</b>	34.42	25.32	20.15
zebra	47.21	55.30	<b>61.51</b>	43.96	41.69	30.93
<b>ORT.</b>	45.35	48.32	<b>54.34</b>	46.49	42.17	38.94

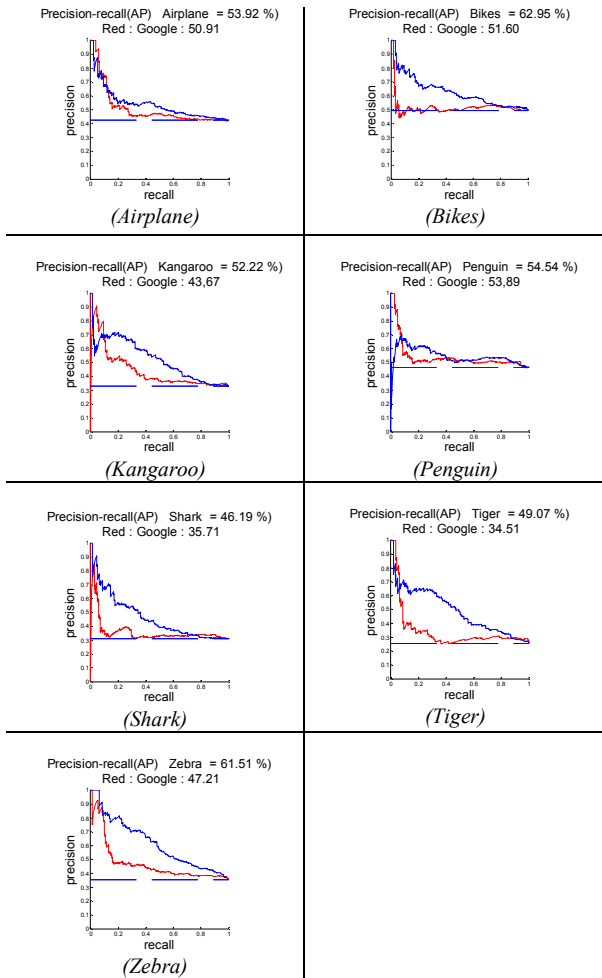
Tablo 1'de de görüldüğü gibi negatif torba boyu ve pozitif torba boyu eşit olduğunda, hem Google arama motoru sonuçlarından daha iyi, hem de değişen pozitif torba boyları arasında genelde en iyi sonuçları elde ediyoruz. Torba boyu  $k_p=10$ 'da en iyi değeri verdiğini ve daha büyüdükçe sonuçların nispeten kötüleşmeye başladığını gözlemlemekteyiz.

Tablo 2'de ise pozitif ve negatif torbalar için eşit örnek sayısı kullanıldığında ( $k_p = k_n$ ) en uygun boyun hangisi olduğunun araştırılması sonuçları verilmiştir. Buna göre küçük torba boyları büyük torba boylarından daha iyi sonuçlar vermektedir. Fakat en küçük torba boyunu kullanmak her zaman iyi olmayabilir, çünkü her ne kadar küçük ve güçlü pozitif torbalar oluşturulsa da,  $k$  değerinin çok küçük seçilmesi durumunda, içinde hiç pozitif görüntü olmayan, fakat pozitif torba etiketi verilen, yanlış etiketli torba sayısı artabilir. Bu nedenle, nispeten daha büyük bir torba boyu ( $k=10$ ) ile daha iyi sonuçlar elde edilebilir.

Tablo 2: Eşit negatif ve pozitif ( $k_p$  ve  $k_n$ ) torba boyları ile sıralamanın precision-recall eğrisi altında kalan alan bakımından karşılaştırılması (yüzde olarak %).

Sorgu	Google	k=5	k=10	k=15	k=20	k=30
airplane	50.91	<b>54.86</b>	53.92	52.74	48.88	49.79
bikes	51.60	61.47	<b>62.95</b>	60.59	60.56	57.04
kangaroo	43.67	49.29	<b>52.22</b>	51.08	51.89	48.40
penguin	53.89	<b>56.03</b>	54.54	50.84	51.93	50.26
shark	35.71	<b>48.72</b>	46.19	46.17	47.82	46.17
tiger	34.51	48.78	<b>49.07</b>	43.57	45.71	42.93
zebra	47.21	57.86	<b>61.51</b>	53.14	55.17	51.48
<b>ORT</b>	45.35	53.86	<b>54.34</b>	51.16	51.70	49.43

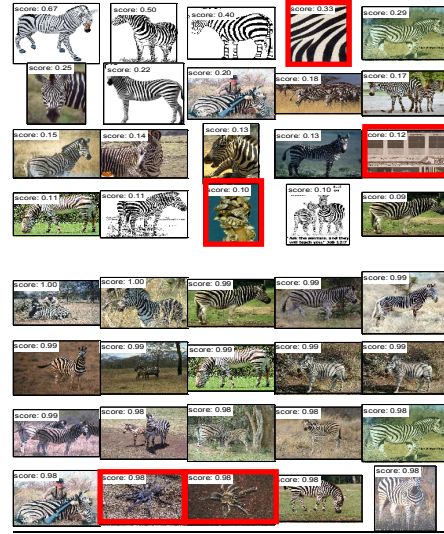
Şekil 4'te her bir sorguya ilişkin kesinlik-geri getirme eğrileri sunulmuştur.



Şekil 4: 7 sorgu için pozitif ve negatif torba boyu  $k_p = k_n = 10$  iken kesinlik-geri getirme (precision-recall) eğrileri. Google arama motorunun eğrisi kırmızı renkte, önerilen yöntemin eğrisini ise mavi renkte gösterilmiştir ve kesikli çizgi rastgele sınıflandırıcıdır.

Şekil 5'te "zebra" sorgusu için dönen ve yeniden sıralanan ilk 20 resim gösterilmiştir. Bu şekillerden de görülebileceği üzere, önerilen yöntem, arama motorunun sonuçlarına göre kesinliği daha yüksek sorgu sıralaması oluşturmakta, resim erişim başarımını arttırmaktadır. Kırmızı

çerçeveli resimler, getirilen yanlış görüntüleri göstermektedir, buna göre Google'dan elde edilen sıralamada bunlar ilk sıralarda yer almakta iken, önerilen yöntemin sıralamasında hem sayıları azalmış hem de ilk başlarda çıkmamaktadırlar.



Şekil 5 : Zebra sorgusu için; ilk 20 görüntü ; üstte Google arama motoru sonucu, altta önerilen yöntemin sıralaması.

## 6. SONUÇ

Bu çalışmada metin tabanlı bir arama motorundan dönen görsel sorguların sonuçlarını, görsel içerikten ve çoklu öğrenme yöntemlerinden yararlanarak yeniden sıralayan bir yaklaşım sunmaktayız. Yöntemimiz, açıktan etiketlemeyi gerektirmeyen çoklu örnekle öğrenme kullanarak örneklerden otomatik torbalar oluşturmak sureti ile sorguyla ilintili görsel yapıyı öğrenmeye çalışmaktadır. Bu şekilde, arama motorundan gelen görüntüler yeniden sıralanmaktadır. Önerilen yöntemimiz, arama motorunun sonuçları ile karşılaştırıldığında, genel olarak ilk sıralarda gelen görüntülerin daha ilgili olduğu görülmektedir. Şu aşamada 7 sorgu kümesi ile denediğimiz yöntemimizi ileride daha geniş veri kümeleri ile denemeyi düşünmekteyiz.

## 7. REFERANSLAR

- Keeler, J.D., Rumelhart, D.E. and Leow, W.K., "Integrated segmentation and recognition of hand-printed numerals." In: NIPS. (1990)
- Dietterich, T.G., Lathrop, R.H. and Lozano-Perez, T., "Solving the multiple-instance problem with axis parallel rectangles", Artificial Intelligence, vol. 89, nos. 1-2, pp. 31-71, 1997.
- Maron. O. and Lozano-Pérez, T., "A framework for multiple-instance learning", Advances in Neural Information Processing Systems 10, pp. 570-576, 1998
- Chen, Y., Bi, J. and Wang, J. Z., "MILES: Multiple-Instance Learning via Embedded Instance Selection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no.12, pp. 1931-1947, 2006.
- Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A. and Freeman, W. T., "Discovering object categories in image collections." In Proceedings of ICCV, 2005.
- Lowe, D. G., "Distinctive image features from scale-invariant key points," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110
- Schroff, F., Criminisi, A. and Zisserman, A., "Harvesting Image Databases from the Web". ICCV 2007