

# TasvirEt: Görüntülerden Otomatik Türkçe Açıklama Oluşturma İçin Bir Denektaş Veri Kümesi

Mesut Erhan Ünal, Begüm Çıtamak, Semih Yağcıoğlu, Aykut Erdem,  
Erkut Erdem, Nazlı İkizler Cinbiş, Ruken Çakıcı



# Haberlerde..



aramak istediğiniz kelimeyi giriniz

## KATEGORİLER

- Tümü
- Türkiye
- Dünya
- Ekonomi
- Spor
- Sağlık
- Yaşam
- Teknoloji
- Emlak
- Sanat
- Otomobil
- Eğitim

## PROGRAMLAR

- Gece Gündüz
- Önce Söz Vardı



### Türk mühendisler görüntüyü Türkçe ses

© 18.04.2016

Facebook, Google ve Microsoft'un üzerinde çalıştığı görüntü betimleme sayesinde görüntüler sesle betimlendi.



# Problem



# Problem



Bir çocuk limonata standında kurabiye yiyor.

Bir limonata standı kurabiye yiyen bir çocuğun yanında.

Küçük bir çocuk kurabiye ve limonata standının yanında kurabiye yiyor.

Bir çocuk limonata standının yanında kurabiye yiyor.

Küçük bir çocuk limonata satarken kurabiye yiyor.

# İlgili Çalışmalar

- Görüntülerden kelime bazlı altyazı getirimi yapılan çalışmalar:
  - Görüntü Etiketleme (örn., Mori vd. 1999, Duygulu vd. 2002)
  - Anlamsal Segmentasyon (örn., Shotton vd. 2008, Tighe ve Lazebnik 2010)
  - İnsanlar, nesneler ya da sahnelerin tanınması (örn. Everingham vd. 2010)
  - Görsel Nitelikler (örn. Farhadi vd. 2009)

# Görüntü Altyazılama Yöntemleri

- Oluşturma Tabanlı Altyazılama (örn., A. Farhadi vd. 2010, G. Kulkarni vd. 2013, M. Mitchell vd. 2012, A. Karpathy vd. 2015)
- Getirim Tabanlı Altyazılama (örn., V. Ordonez vd. 2011, P. Kuznetsova vd. R. Mason 2014, M. Kilickaya vd. 2015)

# Diğer Veri Kümeleri

# SBU Veri Kümesi



Man sits in a rusted car buried in the sand on Waltarere beach



Little girl and her dog in northern Thailand. They both seemed interested in what we were doing



Interior design of modern white and brown living room furniture against white wall with a lamp hanging



Emma in her hat looking super cute

- Flickr adlı fotoğraf paylaşım sitesinde kullanıcıların paylaştığı, nesne, nitelik, eylem ve sahneleri içeren yaklaşık 1,000,000 görüntü, kullanıcı tarafından yazılmış bir açıklamaya sahip. (ortalama 10 kelime).



# Flickr8k Veri kümesi



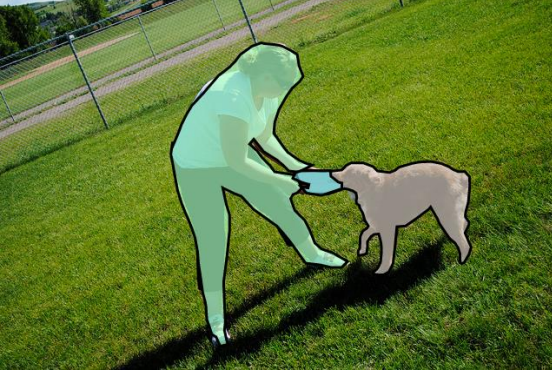
- A black, white, and brown dog is playing with a green and pink ball on sand.
  - A dog is in the sand with a pink and green tennis ball in his mouth.
  - A little dog plays with a colorful ball in the sand.
  - A small terrier plays with a tennis ball on the beach.
  - The white dog, with dark patches, is playing in the sand with a tennis ball in its mouth.
- 
- Flickr isimli fotoğraf paylaşım sitesinde kullanıcıların paylaştığı nesne ve hareket içeren 8,091 resim.
  - Her resim 5 alternatif açıklamaya sahiptir (ortalama 10 kelime uzunluğunda). Amazon Mechanical Turk ile toplanmıştır .

# Flickr30k Veri Kümesi



- Gray haired man in black suit and yellow tie working in a financial environment.
  - A graying man in a suit is perplexed at a business meeting.
  - A businessman in a yellow tie gives a frustrated look.
  - A man in a yellow tie is rubbing the back of his neck.
  - A man with a yellow tie looks concerned.
- 
- Flickr isimli fotoğraf paylaşım sitesinde kullanıcıların paylaştığı nesne ve hareket içeren 31,783 resim.
  - Her resim 5 alternatif açıklamaya sahiptir (ortalama 10 kelime uzunluğunda). Amazon Mechanical Turk ile toplanmıştır .

# MS COCO Veri Kümesi



- a woman playing tug of war with a dog over a white frisbee.
- a woman wrestles a frisbee being held by a dog.
- a dog is biting a frisbee his owner is trying to get
- woman and dog with frisbee on grass near fence.
- a dog is playing with their owner and a frisbee.

- Veri kümesi toplamda 300,000 görüntüden oluşmaktadır.
- Her bir görüntüye ait 5 betimleme bulunmakta, ortalama altyazı 10 kelimeden oluşmaktadır.
- Görüntülerde bulunan nesneler 80 kategoride segmente edilmiştir.

# TasvirEt Veri Kümesi

## Merhaba!

Laboratuvarımız bünyesinde yürütülen bir araştırma projesinde kullanılmak üzere oluşturulacağımız veri kümesinin yaratılmasında, kendisine gösterilen fotoğraflara Türkçe açıklama yazacak ve/veya yazılan açıklamaların kalitesini oylayacak gönüllülere ihtiyaç duyulmaktadır. Veri kümemizi oluşturmada bize yardım etmek isterseniz, aşağıdaki ilgili formlar aracılığı ile sisteme kayıt olabilirsiniz ya da giriş yapabilirsiniz. Proje hakkında detaylı bilgi için [buraya](#) tıklayabilirsiniz. Teşekkürler!

-Proje Grubu

### Yeni Üyelik

**Adınız Soyadınız:**

**E-Posta Adresiniz:**

Aktivasyon iletiniz bu adrese gönderilecektir.

**Şifreniz:**

**Şifreniz (Tekrar):**

**Eğitim Durumunuz:**

---Lütfen Seçiniz---

### Üye Girişi

**E-Posta Adresiniz:**

**Şifreniz:**

### En Çok Açıklama Yazarlar

1. Erkut Erdem (3225 açıklama)  
30/01/2016 7:41 tarihinde
2. Mert Kılıçkaya (2062 açıklama)  
31/01/2016 18:14 tarihinde
3. Semih Yağcıoğlu (2058 açıklama)  
31/01/2016 2:19 tarihinde
4. Aykut Erdem (485 açıklama)  
6/08/2015 7:40 tarihinde
5. Begüm Çıtamak (472 açıklama)  
31/01/2016 22:31 tarihinde
6. Yasin KAVAK (125 açıklama)  
15/04/2015 17:04 tarihinde
7. Burak Kerim Akkuş (123 açıklama)  
9/04/2015 10:55 tarihinde
8. Mesut Erhan Ünal (108 açıklama)  
9/06/2015 17:35 tarihinde
9. Balta Pala (107 açıklama)  
22/03/2015 23:32 tarihinde
10. Kemal GİZMECİLER (83 açıklama)  
10/03/2015 11:29 tarihinde

## Hatırlatmakta Fayda Görüyoruz

Açıklama yazmaya geçmeden önce, yazılacak açıklamalar ile ilgili kısıtları ve iyi bir açıklamanın nasıl olması gerektiğini tekrar hatırlamak, projemize vermiş olduğunuz desteğin kalitesini artırabilir.

Yazacağınız açıklamalar:

- Tek cümleden oluşmalı,
- Yüz kırk (140) karakteri geçmemeli,
- Fotoğraftaki varlıklar ve eylemleri doğru bir şekilde betimlemeli,
- İmlâ kuralları göz önüne alınarak yazılmalıdır.

Örneğin:



· Kırmızı sombrero takmış bir köpek.

**Çok İyi:** Fotoğraftaki tüm ana varlıklar, tam olarak ve uygun bir şekilde açıklanmış.

· Kırmızı şapkalı bir köpek.

**İyi:** Bu şekildeki tamamlanmamış cümleler kabul edilebilir.

· Pembe tasmalı beyaz bir köpek.

**Orta:** Köpeği açıklamış ama şapkadan bahsetmemiş.

· Altın pullarla süslü bir şapka.

**Kötü:** Şapkayı detaylı şekilde açıklamış ancak köpekten bahsetmemiş.

· Köpek

**Çok Kötü:** Yeterli detay yok. Herhangi bir fotoğraftaki, herhangi bir köpek olabilir.

Açıklama Yazmaya Başla

1 / 10



Ormanda kayalardan zıplayan erkek çocuğu

100 karakter kaldı.

Açıklamayı Gönder

Bu Fotoğraflı Atla

# Verinin Toplanması

- Türkçe açıklama sayısı: 2 farklı açıklama görüntü başına
- Ortalama açıklama uzunluğu: 8 kelime
- Bir kere görülen kelime sayısı: 2041 kelime



**Description:** A rollerblader is sliding down a railing.



**Translate:** Bir rollerblader korkuluk aşağı kayıyor



**Açıklama:** Merdiven korkuluklarından kaykayı ile inen bir genç



**Description:** A little girl on a kid swing



**Translate:** Bir çocuk salıncak üzerinde küçük bir kız.

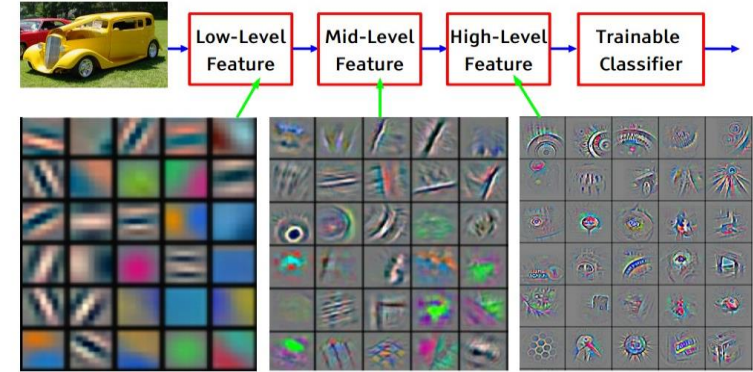
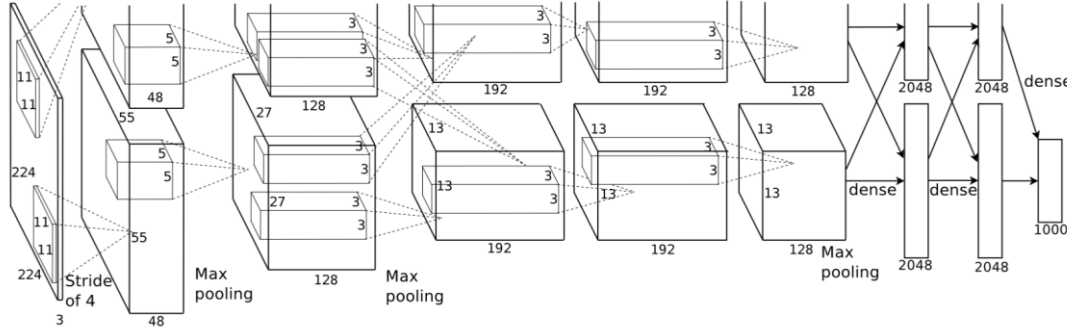


**Açıklama:** Küçük bir kız çocuğu gülerек salıncakta sallanıyor.



# En yakın görüntünün getirilmesi

- Görsel olarak görüntünün tanımının yapılması.
- ImageNet veri kümesi üzerinde eğitilmiş bir derin öğrenme modeli olan 16-katmanlı Evrişimli Yapay Sinir Ağının fc7 aktivasyonları ile 4096 boyutlu vektörler.



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

# En yakın görüntünün getirilmesi



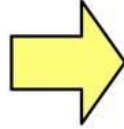
Joulin et al. “Learning visual features from Large Weakly supervised Data”, 2015

# En yakın görüntünün getirilmesi

- K-NN yaklaşımı ile K adet görüntünün getirilmesi.



Aday Görüntü  $I_q$



Görsel Olarak En Yakın Görüntülerin Getirilmesi

$I_1$



C<sub>11</sub>: Karda köpeği ile koşan bir adam.

C<sub>12</sub>: Karda yürüyen biri ve arkasından gelen siyah bir köpek .

$I_2$



C<sub>21</sub>: Karda kasklı turuncu montlu bir çocuk.

C<sub>22</sub>: Karda göz kırpan siyah kasklı oğlan çocuğu.

⋮

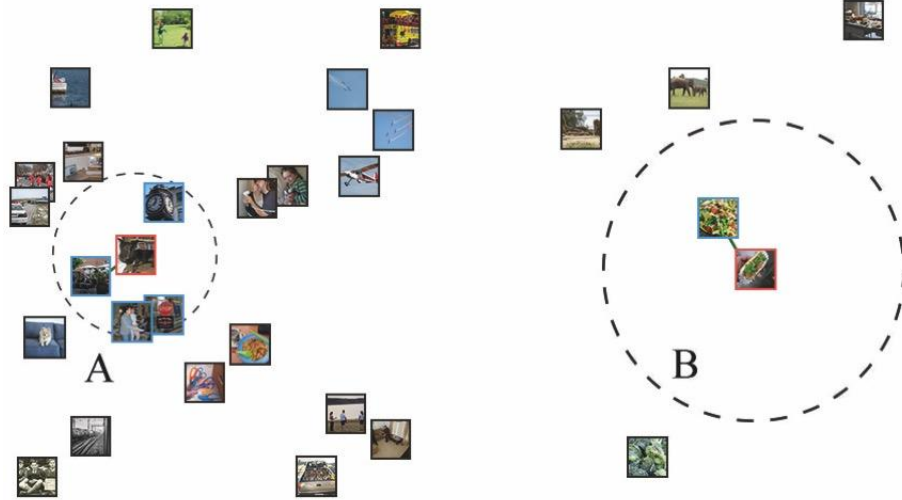
$I_5$



C<sub>51</sub>: Karda kayan adam.

C<sub>52</sub>: Karda yürüyen kırmızı montlu bir adam.

# Uyarlanabilir Komşuluk



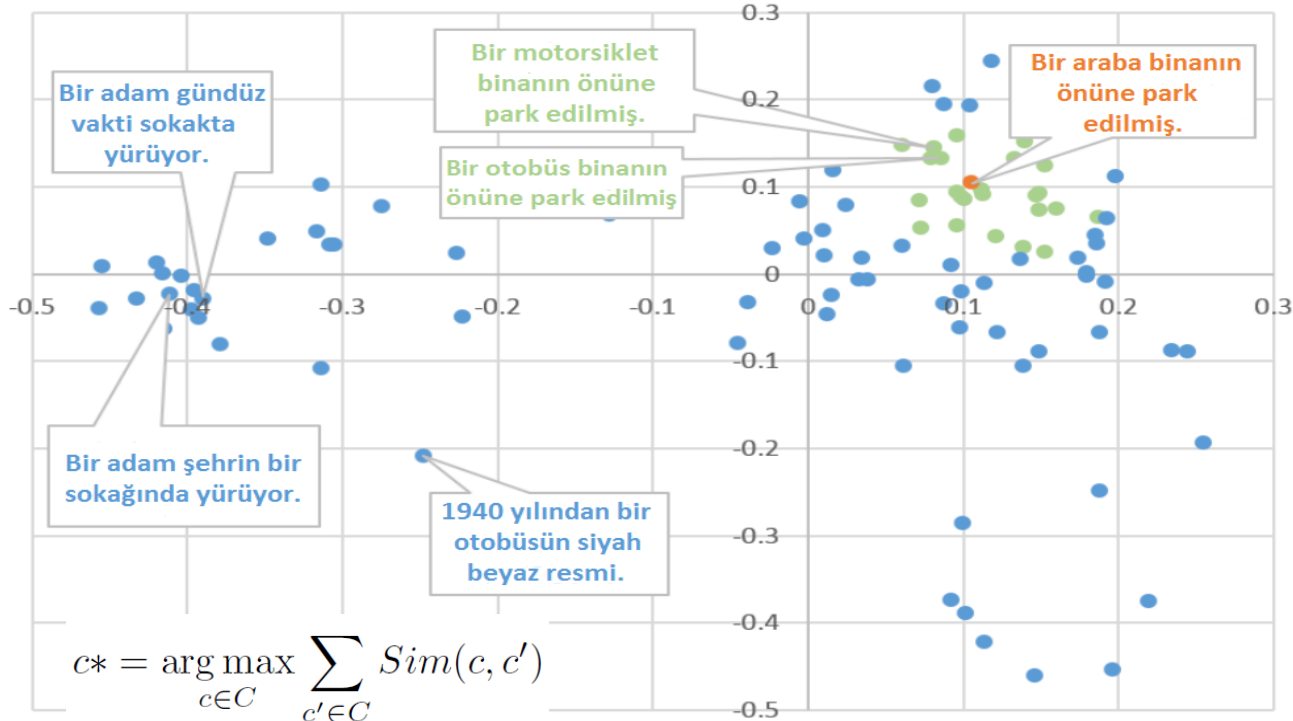
$$\begin{aligned}
 \mathcal{N}(I_q) &= \{(I_i, c_i) \mid \text{dist}(I_q, I_i) \leq (1 + \epsilon) \text{dist}(I_q, I_{\text{closest}}), \\
 I_{\text{closest}} &= \arg \min \text{dist}(I_q, I_i), I_i \in \mathcal{T}\}
 \end{aligned} \tag{1}$$

# Altyazıların Gösterilmesi

- Zemberek ile kelimelerin çekim eklerinin atılması

<b>Altyazı</b>	Kucakladığı	sörf	tahtasıyla	okyanusa	giren	bir	adam
<b>Kökler</b>	kucakla	sörf	tahta	okyanus	gir	bir	adam

# Oybirliği ile Altyazının Bulunması



# BLEU Skorunun Hesaplanması

- Cümle cümle n-gram eşleşmeleri bulunur,
- Her aday cümle için kırılmış n-gram sayısı toplanır,
- Aday cümlelerin n-gram sayısına bölünür.

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n-gram \in C} Count_{clip}(n-gram)}{\sum_{C' \in \{Candidates\}} \sum_{n-gram' \in C'} Count_{clip}(n-gram')}$$

$$Count_{clip} = \min(Count, Max\_Ref\_Count)$$

# Başarılı Sonuçlar





G: Kullanıcı altyazısı

Y1: En yakın görüntünün  
altyazısı

Y2: Oybirliği ile bulunan altyazı

G: Sarı bir köpek merdivenlerden koşarak iniyor.

Y1: Sarı bir köpek çeşmeden su içiyor.

Y2: Çimlerde koşan bir köpek.



G: Kullanıcı altyazısı

Y1: En yakın görüntünün  
altyazısı

Y2: Oybirliği ile bulunan altyazı

G: Bir adam tekne ile su kayağı yapıyor.

Y1: Bir adam sörf yapıyor.

Y2: Azgın dalgaların arasında sörf yapan bir adam.



G: Kullanıcı altyazısı

Y1: En yakın görüntünün  
altyazısı

Y2: Oybirliği ile bulunan altyazı

G: Dalgalarda havalanan bir sörfçü.

Y1: El arabasına oturmuş bir kız çocuğunu çeken patenli bir bayan.

Y2: Dalgada sörfle kayan bir adam.





G: Kullanıcı altyazısı

Y1: En yakın görüntünün  
altyazısı

Y2: Oybirliği ile bulunan altyazı

G: Tenis raketi ile topa vuran kadın tenisçi.

Y1: Bir adam elinde tenis raketi ile koşuyor.

Y2: Tenis oynayan bir kadın.

# Başarısız Sonuçlar



G: Kullanıcı altyazısı  
Y1: En yakın görüntünün  
altyazısı  
Y2: Oybirliği ile bulunan altyazı

G: Miami oyuncusu zenci basketbolcu.  
Y1: Sporcu bir kızın dışardan görüntüsü.  
Y2: Endişeli endişeli bir yerlere bakmakta olan bir kadın.



G: Kullanıcı altyazısı

Y1: En yakın görüntünün  
altyazısı

Y2: Oybirliği ile bulunan altyazı

G: Rugby maçında topu kapma mücadelesi veren oyuncular.

Y1: Elinde top ile rakibine doğru giden rugby oyuncusu ve arkasında takım arkadaşları.

Y2: Bir futbol maçı.

# Test Sonuçları

- 1000 test görüntüsü için  $n=1, 2$  ve  $3$  olduğunda elde edilen sonuçlar:

	BLEU-1	BLEU-2	BLEU-3
Y1	0.211	0.072	0.020
Y2	<b>0.260</b>	<b>0.102</b>	<b>0.034</b>



# Sonuç

- Flickr8k veri kümesinin Türkçe ile zenginleştirilerek TasvirEt veri kümesinin oluşturulması,
- Türkçe altyazılama amacı ile kullanılabilecek iki yöntemin oluşturulması
  - Görsel olarak en yakın görüntünün altyazısının kullanılması
  - Oybirliği ile altyazının bulunması
- Veri kümesindeki altyazıların arttırılması

# Gelecek Çalışmalarımız

- Türkçeye özel cümle temsillerinin oluşturulması
  - Getirme tabanlı bir yaklaşım yerine yaratma tabanlı bir yaklaşım
  - Çok dilli görüntü altyazılama
  - Çok kipli makine çevrimi
- 
- Bu çalışmalar için daha çok veriye ihtiyaç duymaktayız.  
Destek olmak için: <http://tasviret.cs.hacettepe.edu.tr/>

# Teşekkürler!