

# Sıralama ve aktarma tabanlı yaklaşımalar



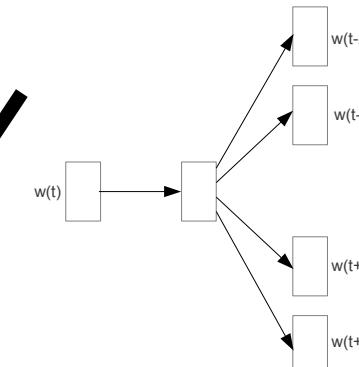
# Görüntü altyazılıama

- Sıralama tabanlı yaklaşımalar
- Aktarma tabanlı yaklaşımalar
- Oluşturma tabanlı yaklaşımalar

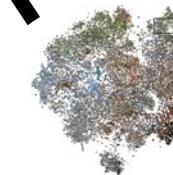
INPUT

PROJECTION

OUTPUT



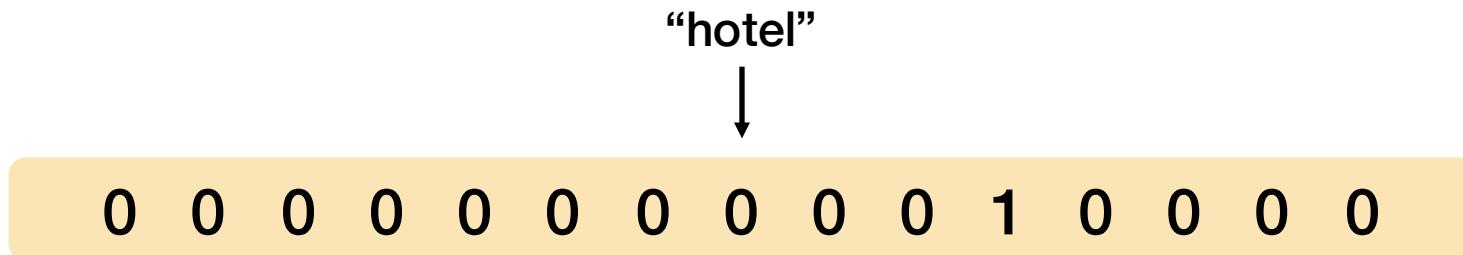
**sinirsel sözcük  
gömmeleri  
(neural word  
embeddings)**



derin görsel öznitelikler

# Kelimelerin vektörel temsilleri

- Doğal dil işlemede kural tabanlı veya istatistiksel yaklaşımların çoğu sözcükleri atomik semboller olarak kabul eder.
- Böyle bir temsil, vektör uzayı anlamında sözcüklerin aşırı seyrek vektörler (tek bir 1 ve diğerleri hep 0) ile ifade edilmelerini doğurur:



- Bu tür bir temsil **tek-sıcak (one-hot)** olarak anılır.

# Tek-sıcak temsili kullanımının sakıncaları

- Sözcüklerin vektörel uzunlukları sözlük boyutuna eşittir.
  - Vektör boyutu sözlük boyutu arttıkça artmaktadır.  
20K (konuşma) – 50K (Penn Treebank) – 500K (büyükçe bir sözlük) – 13M (Google 1T).
- Vektör temsilleri birbirine dikgen (*orthogonal*) oldukları için benzerlik kavramı doğal bir şekilde tanımlanamamaktadır.

	T														
“motel”	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
“hotel”	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

= 0

# Dağıtılmış benzerliğe dayalı temsiller

- Bir sözcük etrafındaki sözcükler üzerinden temsil edildiğinde o sözcüğe dair bir sürü bilgi edinilebilir.
- “You shall know a word by the company it keeps”  
(J. R. Firth 1957:11)
- Modern dil işlemenin en başarılı düşünülerden biridir.



government debt problems turning into banking crises as has happened in  
saying that Europe needs unified banking regulation to replace the hodgepodge

“banking” sözcüğünün temsili bu komşu sözcüklere dayanmaktadır

# Dağıtılmış benzerliğe dayalı temsiller

He filled the **wampimuk**, passed it around and we all drunk some  
(**vampimuk** doldurdu, elden ele dolaştırdı ve hepimiz biraz içtik.)

We found a little, hairy **wampimuk** sleeping behind the tree  
(Bir ağacın arkasında uyuyan, küçük, tüylü bir **vampimuk** bulduk.)

# Dağıtılmış benzerliğe dayalı temsiller

he curtains open and the moon shining in on the barely  
ars and the cold , close moon " . And neither of the w  
rough the night with the moon shining so brightly , it  
made in the light of the moon . It all boils down , wr  
surely under a crescent moon , thrilled by ice-white  
sun , the seasons of the moon ? Home , alone , Jay pla  
m is dazzling snow , the moon has risen full and cold  
un and the temple of the moon , driving out of the hug  
in the dark and now the moon rises , full and amber a  
bird on the shape of the moon over the trees in front  
But I could n't see the moon or the stars , only the  
rning , with a sliver of moon hanging among the stars  
they love the sun , the moon and the stars . None of  
the light of an enormous moon . The plash of flowing w  
man 's first step on the moon ; various exhibits , aer  
the inevitable piece of moon rock . Housing The Airsh  
oud obscured part of the moon . The Allied guns behind

A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge  
[Landauer and Dumais'97]

From frequency to meaning: Vector space models of semantics [Turney ve Pantel'10]

# Dağıtılmış sözcük gömmelerinin yapay sinir ağları ile öğrenilmesi

- Önerilen bütün modeller bir sözcüğü etrafındaki diğer sözcüklerin oluşturduğu bağlam üstünden tahmin etmeye dayanmaktadır:

$$\operatorname{argmax}_w w \cdot ((w_{j-1} + w_{j+1}) / 2)$$

Birim norm  
vektörleri

- Bir kayıp fonksiyonu, mesela  $J(\theta) = 1 - w_j \cdot ((w_{j-1} + w_{j+1}) / 2)$
- Büyük bir dil bütüncesi (*corpus*) içinde çok sayıda örneğe bakılır.
- Stokastik Gradyan iniş yöntemi ile enküçültme:

$$\theta^{new} = \theta^{old} - \alpha \nabla_{\theta} J(\theta)$$

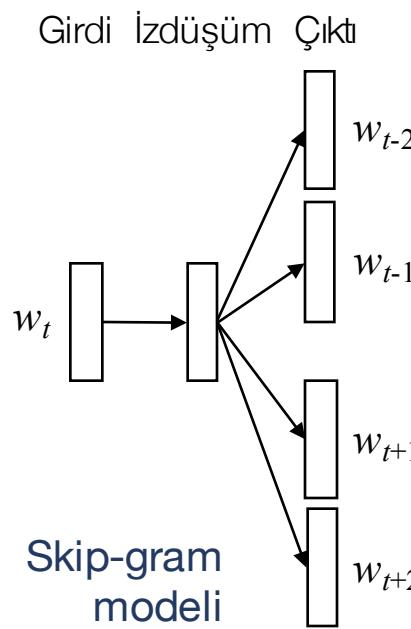
Distributed representations of words and phrases and their compositionality [Mikolov vd.'13]

GloVe: Global Vectors for Word Representation [Pennington vd.'14]

Yansı: Christopher D. Manning

# Word2Vec Skip-gram modeli

- Amaç fonksiyonu:** Merkez sözcük verildiğinde ilgili bağlam sözcüklerinin log olasılık değerinin enbüyültülmesi



$$J(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log p(w_{t+j} | w_t)$$

$$p(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w=1}^W \exp(u_w^T v_c)}$$

$$\theta = \begin{bmatrix} v_{aardvark} \\ v_a \\ \vdots \\ v_{zebra} \\ u_{aardvark} \\ u_a \\ \vdots \\ u_{zebra} \end{bmatrix} \in \mathbb{R}^{2dV}$$

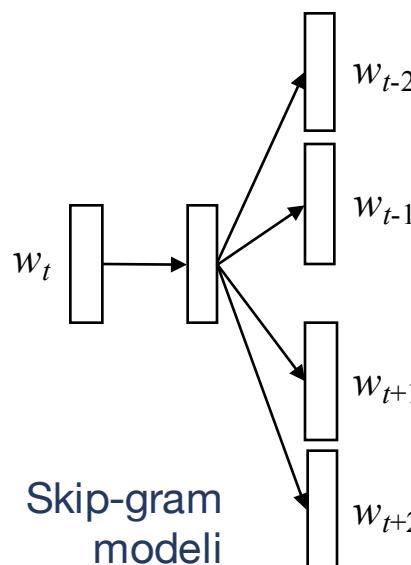
**Her sözcüğün iki farklı  
vektörel temsili vardır!**

$$x_{final} = u + v$$

# Word2Vec Skip-gram modeli

- Amaç fonksiyonu:** Merkez sözcük verildiğinde ilgili bağlam sözcüklerinin log olasılık değerinin enbüyükültülmesi

Girdi İzdüşüm Çıktı



$$J(\theta) = \frac{1}{T} \sum_{t=1}^T J_t(\theta)$$

**Negatif örneklem:**

$$J_t(\theta) = \log \sigma(u_o^T v_c) + \sum_{j \sim P(w)} [\log \sigma(-u_j^T v_c)]$$

**başlangıç sözcüklerinin olasılığı  
yüksek olurken rastgele seçilen  
bir sözcüğün olasılığı düşük olsun**

$$P(w) = U(w)^{3/4} / Z$$

*Daha az gözlemlenen  
sözcükler daha sık seçilir*

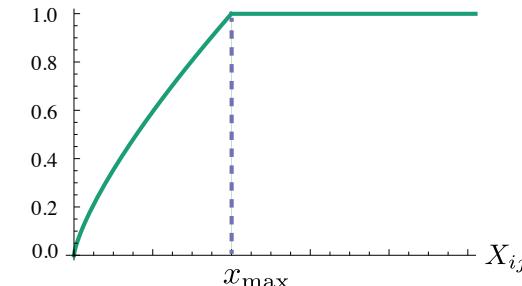
# GloVe modeli

- **Amaç fonksiyonu:** Word2vec Skip-gram modeline kıyasla eğitimi daha kolay bir fonksiyon:

$$J(\theta) = \frac{1}{2} \sum_{i,j=1}^W f(P_{ij}) (u_i^T v_j - \log P_{ij})^2$$

$$f(x) = \begin{cases} (x/x_{\max})^\alpha & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$$

$P_{ij} = P(j|i)$  : sözcük  $j$ 'nin sözcük  $i$ 'nin bağlamında görülme olasılığı



$$\alpha = 3/4$$

$$x_{\max} = 100$$

# Dağıtılmış sözcük gömmeleri anlamsal benzerliği yakalayabilir

- “**frog**” sözcüğüne en yakın kelimeler
  1. frogs
  2. toad
  3. litoria
  4. leptodactylidae
  5. rana
  6. lizard
  7. eleutherodactylus



“litoria”



“leptodactylidae”

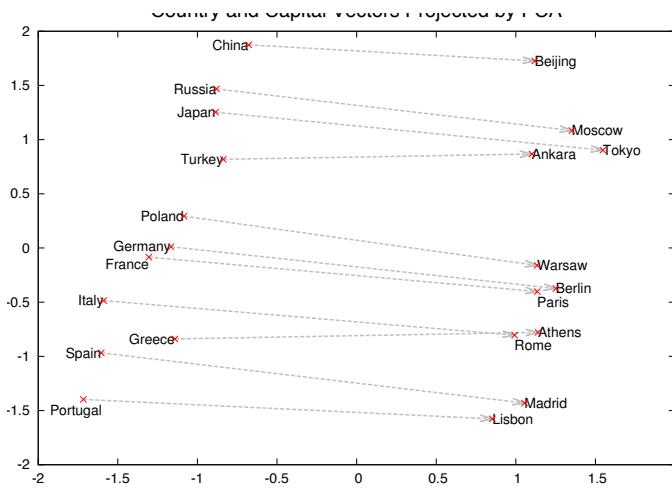


“rana”



“eleutherodactylus”

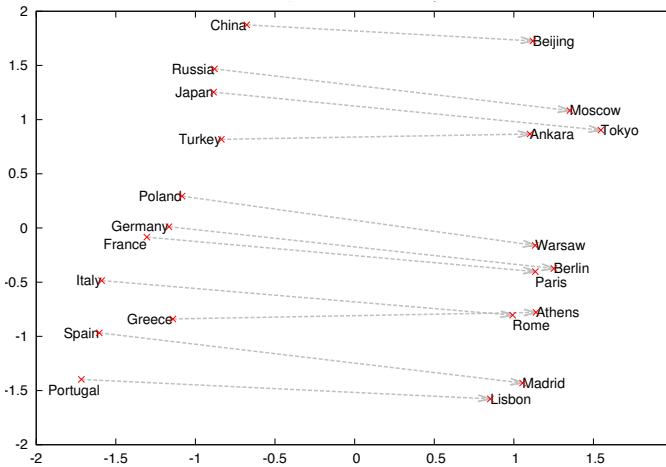
# ve hatta bazı diğer dilbilimsel düzenlilikleri...



$$x_{Paris} - x_{France} + x_{Italy} \approx x_{Rome}$$

Ülkeler ve başkentleri

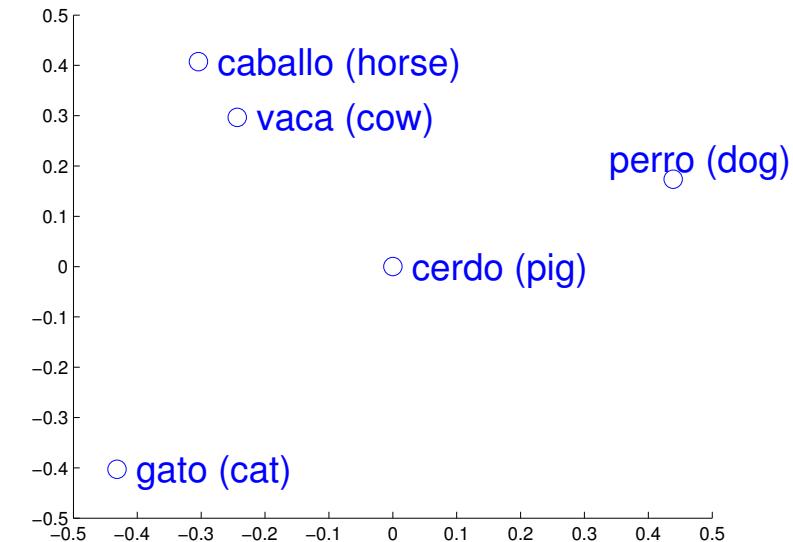
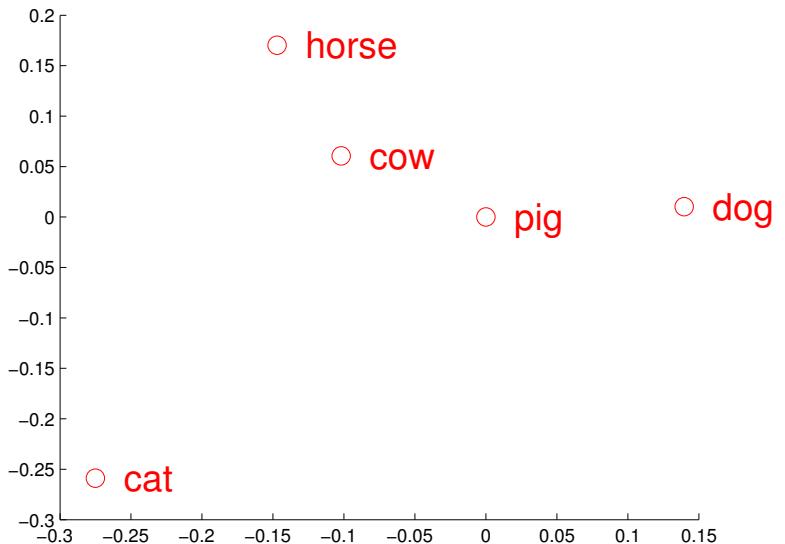
# ve hatta bazı diğer dilbilimsel düzenlilikleri...



Ülkeler ve başkentleri

Expression	Nearest token
Paris - France + Italy	Rome
bigger - big + cold	colder
sushi - Japan + Germany	bratwurst
Cu - copper + gold	Au
Windows - Microsoft + Google	Android
Montreal Canadiens - Montreal + Toronto	Toronto Maple Leafs

# ve hatta bazı diğer dilbilimsel düzenlilikleri...

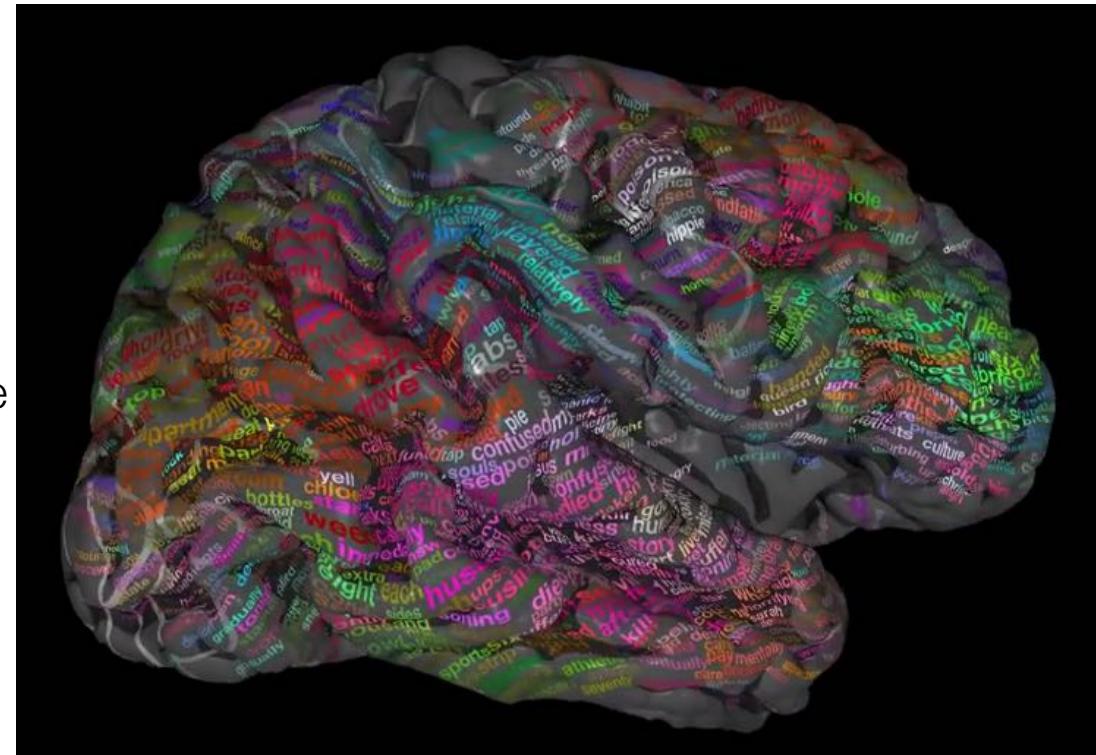


5 farklı hayvanın dağıtılmış sözcük vektörleri (solda: İngilizce, sağda: İspanyolca).

Peki ya insan beyini  
nasıl bir temsil  
kullanıyor?

- Yakın zamanlı bir deneyde bir grup deneğe çeşitli öyküler dinletilmiş ve bu sırada voksel yolu fMRI verisi toplanmıştır.
  - Bu toplanan verinin analizi, sözcüklerin serebral kortekste anlamsal ilişkilere göre diziliyor olabileceğini işaret etmektedir.

<http://gallantlab.org/huth2016/>



<https://www.youtube.com/watch?v=k61nJkx5aDQ>

Natural speech reveals the semantic maps that tile human cerebral cortex [Huth vd., 2016]

# Dağıtılmış temsiller ilginç durumları da ortaya çıkarabilir

## Extreme *she* occupations

- |                 |                       |                        |
|-----------------|-----------------------|------------------------|
| 1. homemaker    | 2. nurse              | 3. receptionist        |
| 4. librarian    | 5. socialite          | 6. hairdresser         |
| 7. nanny        | 8. bookkeeper         | 9. stylist             |
| 10. housekeeper | 11. interior designer | 12. guidance counselor |

## Extreme *he* occupations

- |                |                   |                |
|----------------|-------------------|----------------|
| 1. maestro     | 2. skipper        | 3. protege     |
| 4. philosopher | 5. captain        | 6. architect   |
| 7. financier   | 8. warrior        | 9. broadcaster |
| 10. magician   | 11. fighter pilot | 12. boss       |

Extreme occupations as projected on to  
the *she–he* gender direction

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings [Bolukbasi vd.'16]

# Dağıtılmış temsiller ilginç durumları da ortaya çıkarabilir

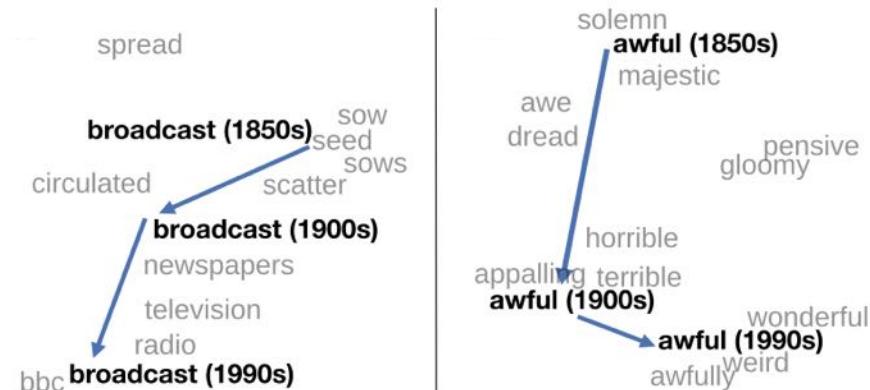
## Extreme *she* occupations

- |                 |                       |                        |
|-----------------|-----------------------|------------------------|
| 1. homemaker    | 2. nurse              | 3. receptionist        |
| 4. librarian    | 5. socialite          | 6. hairdresser         |
| 7. nanny        | 8. bookkeeper         | 9. stylist             |
| 10. housekeeper | 11. interior designer | 12. guidance counselor |

## Extreme *he* occupations

- |                |                   |                |
|----------------|-------------------|----------------|
| 1. maestro     | 2. skipper        | 3. protege     |
| 4. philosopher | 5. captain        | 6. architect   |
| 7. financier   | 8. warrior        | 9. broadcaster |
| 10. magician   | 11. fighter pilot | 12. boss       |

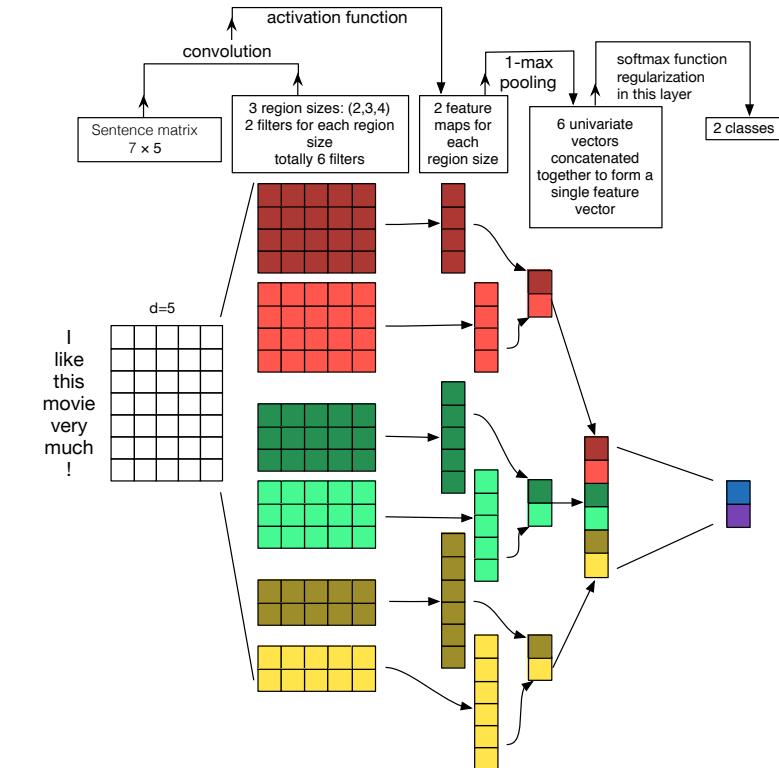
Extreme occupations as projected on to  
the *she–he* gender direction



Visualizing changes in word meaning

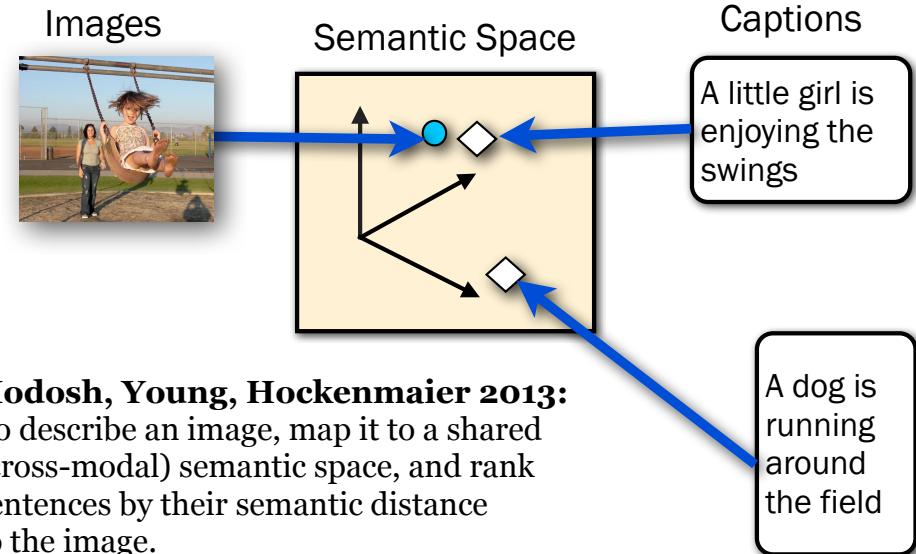
# Bir uygulama örneği: Evrişimsel Sinir Ağları ile Cümle Sınıflandırma

- Verilen bir cümle bir görüntüümüş gibi bir matris ile ifade edilebilir
- $s \times d$  boyutlu matris
  - $s$ : verilen cümlenin uzunluğu
  - $b$ : dağıtılmış sözcük gömmelerinin boyutu
  - Gerektiğinde sıfır doldurma (*zero padding*) stratejisi uygulanır
- Doğrusal filtreler
- Aktivasyon fonksiyonları (ReLU, vb.)
- Toparlama (maksimum, ortalama)

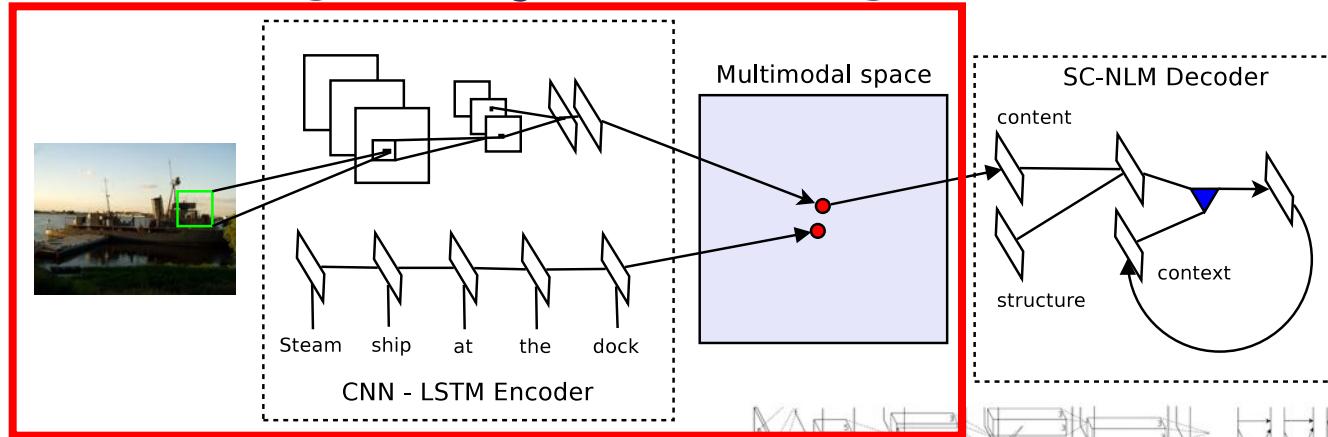


# Sıralama tabanlı görüntü altıyzılama

- Hem görsel hem metinsel veri için ortak bir anlamsal gömme öğrenme
  - Canonical Correlation Analysis (CCA) veya türevi yöntemlere dayalı modeller
  - Derin öğrenme tabanlı modeller
- Bir görüntü girdi olarak verildiğinde bu ortak uzay içinde kipler arası getirme (*cross-modal retrieval*) gerçekleştirilebilir.



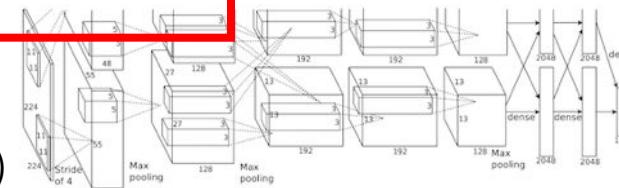
# Çokkipli dağıtılmış temsil öğrenme



- Görüntü ve cümle temsilleri
  - **görüntü**: derin bir evrişimsel ağ (CNN)
  - **cümle**: uzun kısa süreli yinelenen ağ (LSTM)

<https://github.com/ryankiros/visual-semantic-embedding>

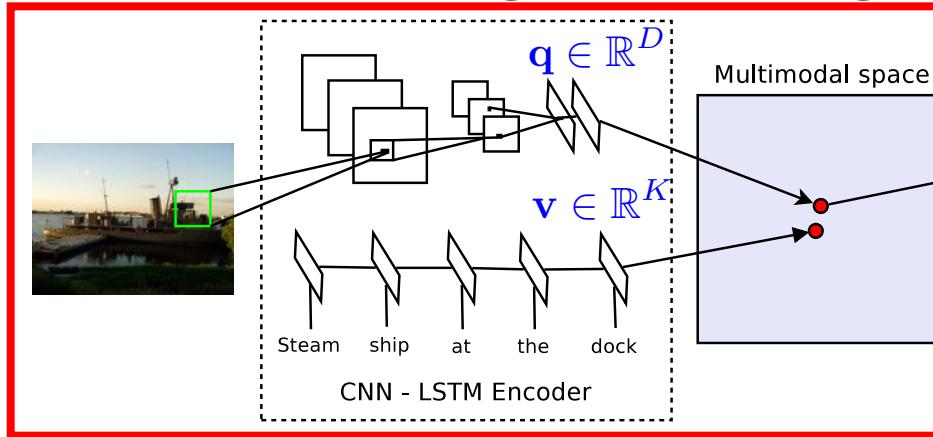
Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models [Kiros vd.'15]



$$\begin{aligned}
 \mathbf{I}_t &= \sigma(\mathbf{X}_t \cdot \mathbf{W}_{xi} + \mathbf{M}_{t-1} \cdot \mathbf{W}_{hi} + \mathbf{C}_{t-1} \cdot \mathbf{W}_{ci} + \mathbf{b}_i) \\
 \mathbf{F}_t &= \sigma(\mathbf{X}_t \cdot \mathbf{W}_{xf} + \mathbf{M}_{t-1} \cdot \mathbf{W}_{hf} + \mathbf{C}_{t-1} \cdot \mathbf{W}_{cf} + \mathbf{b}_f) \\
 \mathbf{C}_t &= \mathbf{F}_t \bullet \mathbf{C}_{t-1} + \mathbf{I}_t \bullet \tanh(\mathbf{X}_t \cdot \mathbf{W}_{xc} + \mathbf{M}_{t-1} \cdot \mathbf{W}_{hc} + \mathbf{b}_c) \\
 \mathbf{O}_t &= \sigma(\mathbf{X}_t \cdot \mathbf{W}_{xo} + \mathbf{M}_{t-1} \cdot \mathbf{W}_{ho} + \mathbf{C}_t \cdot \mathbf{W}_{co} + \mathbf{b}_o) \\
 \mathbf{M}_t &= \mathbf{O}_t \bullet \tanh(\mathbf{C}_t)
 \end{aligned}$$

*Detaylar Nazlı'nın sunumunda!*

# Çokkipli veriden gömme öğrenme



$\mathbf{x} = \mathbf{W}_I \cdot \mathbf{q} \in \mathbb{R}^K$  : görüntü gömmesi

$\mathbf{v} \in \mathbb{R}^K$  : cümle gömmesi

*başlangıçta word2vec  
vektörleri (K=300)*

$s(\mathbf{x}, \mathbf{v}) = \mathbf{x} \cdot \mathbf{v}$  : puanlama fonksiyonu

- **Kayıp fonksiyonu:** İkili sıralama kaybı

$$J(\theta) = \min_{\theta} \sum_{\mathbf{x}} \sum_k \max\{0, \alpha - s(\mathbf{x}, \mathbf{v}) + s(\mathbf{x}, \mathbf{v}_k)\} + \sum_{\mathbf{v}} \sum_k \max\{0, \alpha - s(\mathbf{v}, \mathbf{x}) + s(\mathbf{v}, \mathbf{x}_k)\}$$

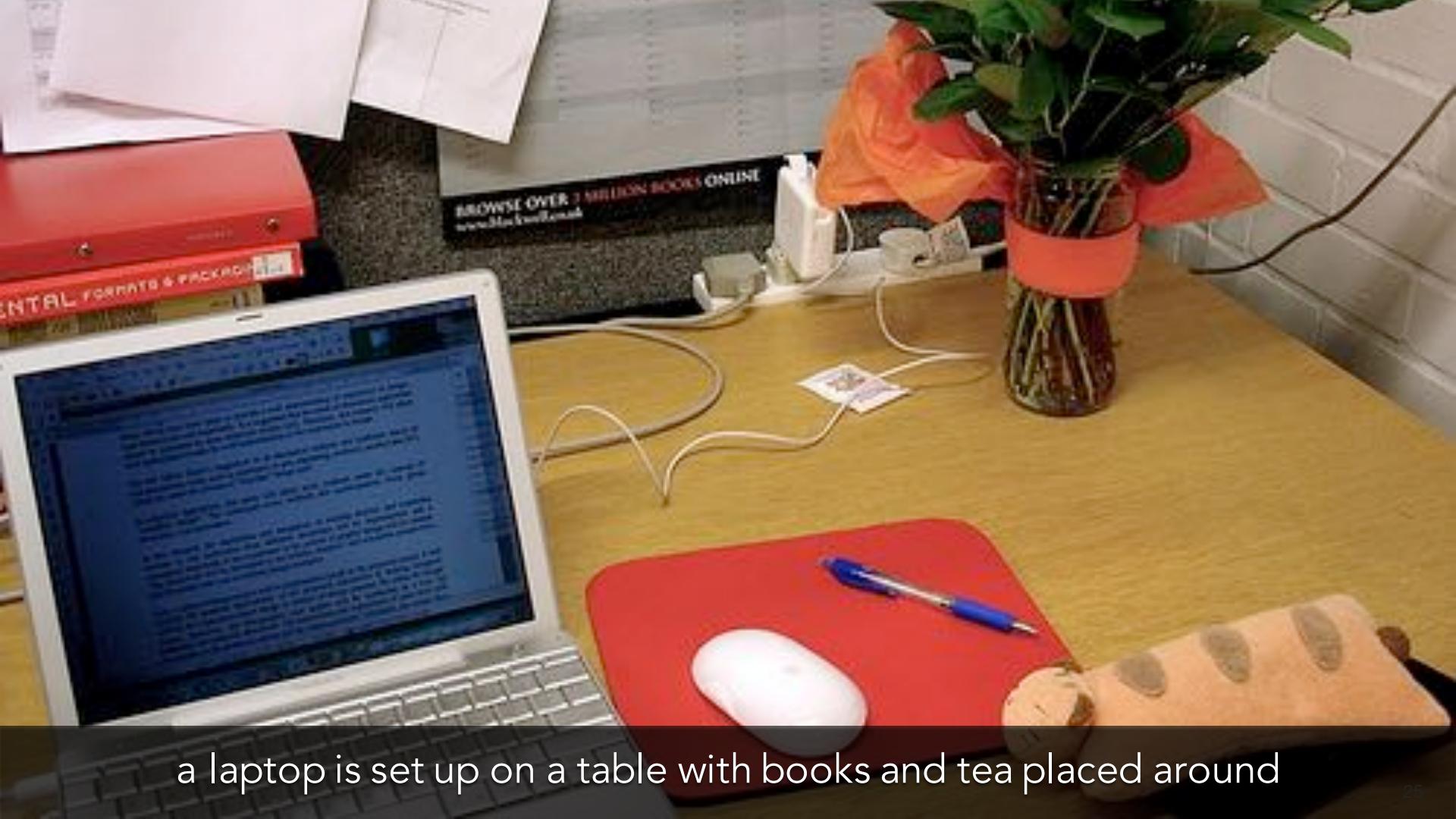
$\mathbf{v}_k$  : karşıtsal (doğru olmayan) bir açıklama



up close view of very pretty bird near the water



a bike that ' s leaning against a cement wall



a laptop is set up on a table with books and tea placed around



a couple of men fishing from a beach by a dog



one zebra laying down in the dirt outside

# Çok kipli dilbilimsel düzenlilikler

Yakın görüntüler



- dog + cat =



- cat + dog =



- plane + bird =



- man + woman =



(a) Basit örnekler

# Çok kipli dilbilimsel düzenlilikler

Yakın görüntüler



- blue + red =



- blue + yellow =



- yellow + red =



- white + red =



(b) Renkler

# Çok kipli dilbilimsel düzenlilikler

Yakın görüntüler



- day + night =



- flying + sailing =



- bowl + box =



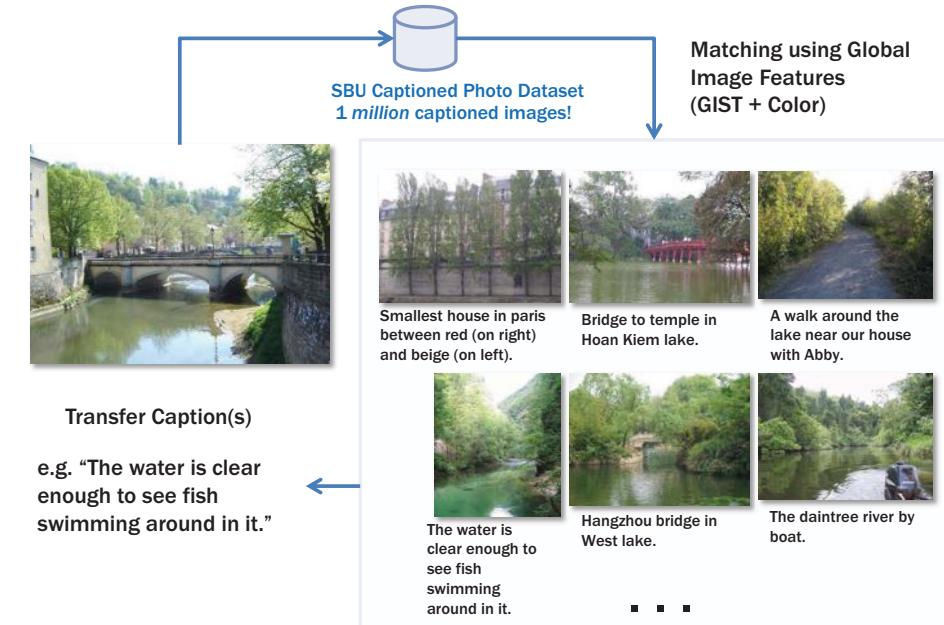
- box + bowl =



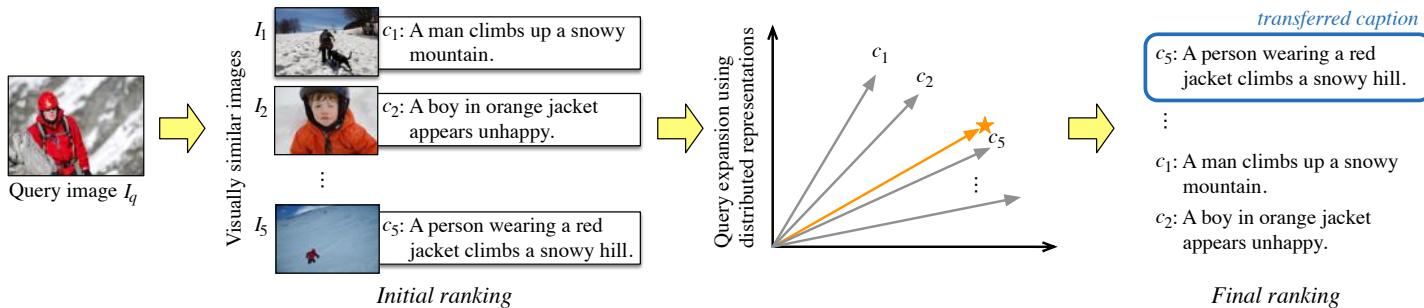
(c) Görüntü yapısı

# Aktarma tabanlı görüntü altyazılılama

- Girdi görüntüsüne benzer görüntüler **büyük bir görüntü kümesi** içinde aranır.
- Benzer görüntüler içinden en ilgili olan görüntünün açıklaması girdi görüntüsünü açıklaması olarak kullanılır.



# Sorgu Genişletme ile Altyazı Aktarma



- Üç aşamalı bir açıklama aktarma yaklaşımı

1. Görsel getirme (*visual retrieval*)
2. Sorgu genişletme (*query expansion*)
3. Tekrar sıralama (*re-ranking*)

} görüp birimi  
} dil birimi

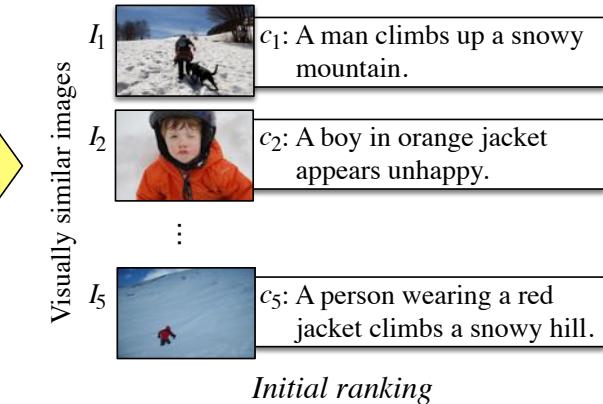
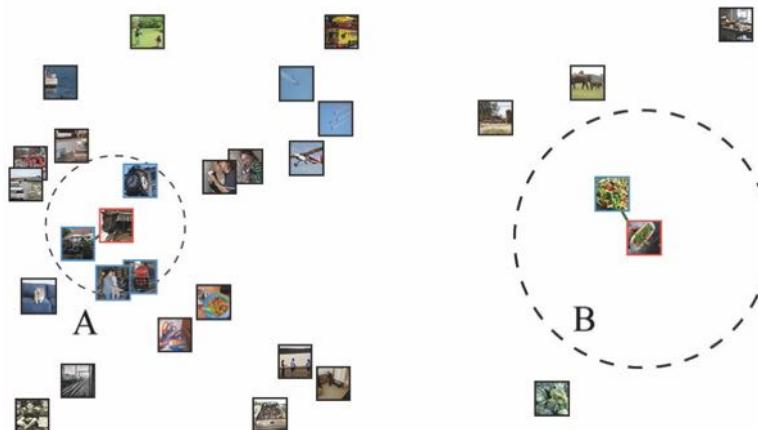


Semih Yagcioglu, Erkut Erdem, Aykut Erdem, Ruket Cakici  
 A Distributed Representation Based Query Expansion Approach for Image Captioning  
 ACL 2015

<https://github.com/semihyagcioglu/image-captioning>

# Görsel getirme

- AlexNet (fc7) öznitelikleri
- Uyarlamalı komşu seçimi

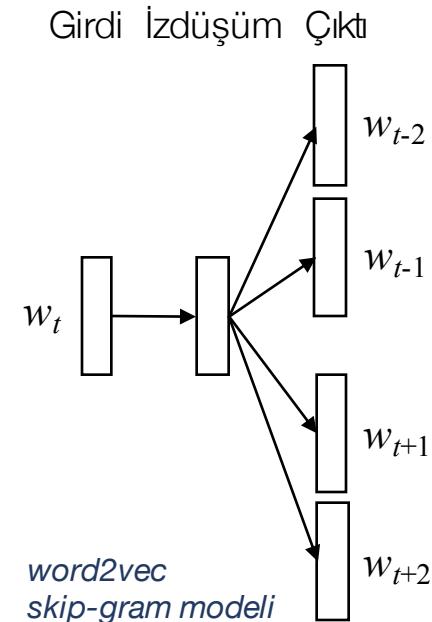


$$\mathcal{N}(I_q) = \{(I_i, c_i) \mid dist(I_q, I_i) \leq (1 + \epsilon)dist(I_q, I_{closest})\}$$

$$I_{closest} = \arg \min dist(I_q, I_i), I_i \in \mathcal{T}$$
(1)

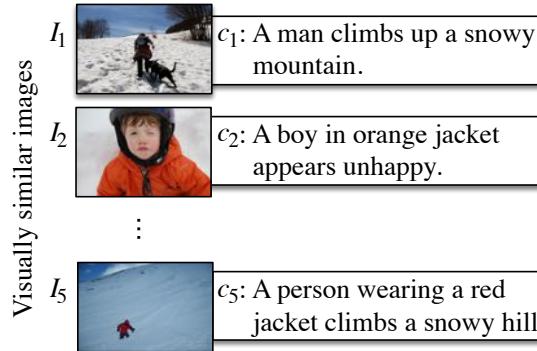
# Cümle temsili

- Dağıtılmış sözcük gömmeleri
  - word2vec model [Mikolov vd., '13]
  - GloVe model [Pennington vd., '14]
- $K = 500$
- MS COCO altyazı bütüncesи (617K)
- Cümle temsili:** Sözcük vektörleri toplamı

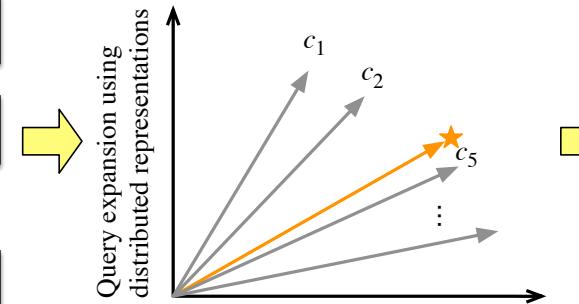


A Comparison of Vector-based Representations for Semantic Composition [Blacoe ve Lapata'12]  
 Distributed representations of words and phrases and their compositionality [Mikolov vd.'13]  
 GloVe: Global Vectors for Word Representation [Pennington vd.'14]

# Sorgu genişletme



Initial ranking



transferred caption

$c_5$ : A person wearing a red jacket climbs a snowy hill.

$\vdots$

$c_1$ : A man climbs up a snowy mountain.

$c_2$ : A boy in orange jacket appears unhappy.

Final ranking

- Bilgi getirmede (*information retrieval*) sıkça kullanılan bir kavram
- Girdi sorgusu getirilen görüntülerin dağıtılmış cümle temsillerinin ağırlıklı ortalaması ile değiştirilir.
- Açıklamalar gömme uzayında tekrar sıralanır

$$q = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M sim(I_q, I_i) \cdot c_i^j$$



a man in a black shirt and his little girl wearing orange are sharing a treat



a construction crew in orange vests working near train tracks



a white cat is sitting in a bathroom sink



a boy is holding a dog that is wearing a hat



a man wearing a santa hat holding  
a dog posing for a picture



a boy is holding a dog that is wearing a hat

# Tasvir Et: Türkçe Görüntü Altyazılıma

- Türkçe görüntü altyazılıma için önerilen ilk veri kümeleri
  - Flickr8K veri kümeleri için Türkçe açıklamalar
  - İngilizce açıklamalardan bağımsız olarak kitle-kaynaklı (*crowd-sourced*) açıklamalar toplandı
- Türkçe altyazılıma için aktarma tabanlı temel (*baseline*) bir yöntem



**SIU 2016 Alper Atalay  
En İyi Bildiri Ödülü**

# TasvirEt Veri Kümesi

<http://tasviret.cs.hacettepe.edu.tr>



## Merhaba!

Laboratuvarımız bünyesinde yürütülen bir araştırma projesinde kullanılmak üzere oluşturulacakımız veri kümесinin 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 olabilir ya da giriş yapabilirsiniz. Proje hakkında detaylı bilgi için [buraya](#) tıklayabilirsiniz. Teşekkürler!

-Proje Grubu

**Yeni Üyelik**

**Adınız Soyadınız:**

**E-Posta Adresiniz:**   
Aktivasyon iletiniz bu adres'e gönderilecektir.

**Şifreniz:**

**Şifreniz (Tekrar):**

**Eğitim Durumunuz:**

**Kayıt Ol** **Aktivasyon İletim Gelmedi**

**Üye Girişi**

**E-Posta Adresiniz:**

**Şifreniz:**

**Giriş** **Şifremi Unuttum**

- En Çok Açıklama Yazanlar**
1. Erkut Erdem ([3544 açıklama](#))  
14/07/2016 9:42 tarihinde
  2. Semih Yagcioglu ([2800 açıklama](#))  
30/06/2016 2:02 tarihinde
  3. Mert Kilickaya ([2244 açıklama](#))  
7/03/2016 19:30 tarihinde
  4. Begüm Çatamak ([756 açıklama](#))  
5/09/2016 7:05 tarihinde
  5. Aykut Erdem ([528 açıklama](#))  
23/04/2016 21:15 tarihinde
  6. Mesut Erhan Ünal ([330 açıklama](#))  
26/05/2016 0:58 tarihinde
  7. Yasin KAVAK ([125 açıklama](#))  
15/04/2015 17:04 tarihinde
  8. Burak Kerim Akkus ([123 açıklama](#))  
9/04/2015 10:55 tarihinde
  9. Balta Pala ([107 açıklama](#))  
22/03/2015 23:32 tarihinde
  10. Kemal CIZMECILER ([83 açıklama](#))  
10/03/2015 11:29 tarihinde



HACETTEPE  
UNIVERSITY  
COMPUTER  
VISION LAB

## Merhaba!

Laboratuvarımız bünyesinde yürütülen bir araştırma yarattılmasında, kendisine gösterilen fotoğraflara Türk oylayacak gönüllülere ihtiyaç duyulmaktadır. Veri ilgili formlar aracılığı ile sisteme kayıt olabilir ya da tıklayabilirsiniz. Teşekkürler!

-Proje Grubu

**Yeni Üyelik**

**Adınız Soyadınız:**

**E-Posta Adresiniz:**

Aktivasyon İletimi bu adres'e gönderilecektir.

**Şifreniz:**

**Şifreniz (Tekrar):**

**Eğitim Durumunuz:**  Lütfen Seçiniz

**Kayıt Ol** **Aktivasyon İletim Gelmedi**

**Üye Girişи**

**E-Posta Adresi:**

**Şifreniz:**

**Giriş** **Sıfırla**



HACETTEPE  
UNIVERSITY  
COMPUTER  
VISION LAB

## Hatırlatmakta Fayda Görüyoruz

Açıklama yazmaya geçmeden önce, yapılacak 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 kuraları 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 şapkadан bahsetmemiş.

- Altın pullarla süslü bir şapka.

**Kötü:** Şapkayı detaylı şekilde açıklamış ancak köpektan bahsetmemiştir.

- Köpek

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

**Açıklama Yazmaya Başla**



HACETTEPE  
UNIVERSITY  
COMPUTER  
VISION LAB

## Merhaba!

Laboratuvarımız bünyesinde yürütülen bir araştırma yarattılmasında, kendisine gösterilen fotoğraflara Türk oylayacak gönüllülere ihtiyaç duyulmaktadır. Veri ilgili formlar aracılığı ile sisteme kayıt olabilir ya da tıklayabilirsiniz. Teşekkürler!

-Proje Grubu

**Yeni Üyelik**

**Adınız Soyadınız:**

**E-Posta Adresiniz:**

Aktivasyon İletimi bu adres'e gönderilecektir.

**Şifreniz:**

**Şifreniz (Tekrar):**

**Eğitim Durumunuz:**  Lütfen Seçiniz...

**Üye Girişи**

**E-Posta Adresi:**

**Şifreniz:**



HACETTEPE  
UNIVERSITY  
COMPUTER  
VISION LAB

## Hatırlatmakta Fayda Görüyoru

Açıklama yazmaya geçmeden önce, yapılacak açıklığı gerektiği tekrar hatırlamak, projemize vermiş olduğu

Yazacağınız açıklamalar:

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

**Örneğin:**



- Kirmizi sombra
- **Cek Iyi:** Fotoğraf
- Kirmizi şapka
- **Iyi:** Bu şekilde
- Pembe tasması
- **Orta:** Köpeği a
- Altın pullarla
- **Kötü:** Şapkayı
- Köpek
- **Cek Kötü:** Yet



HACETTEPE  
UNIVERSITY  
COMPUTER  
VISION LAB

1 / 10



Açıklamanızı buraya yazınız...

140 karakter kaldı.

# Verinin Toplanması

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

Google Translate  
çözüm değil!



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

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

**Google:** Bir rollerblader korkuluk aşağı kayıyor

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

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

**Google:** Bir çocuk salıncak üzerinde küçük bir kız

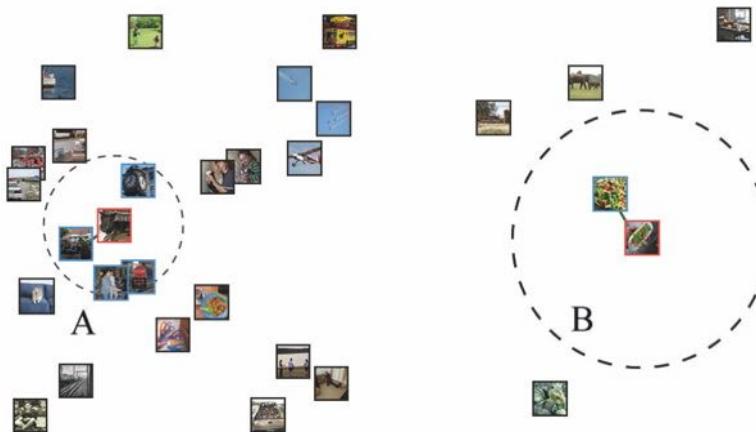
---

# Oybırlığı ile Türkçe Altyazı Aktarma

1. Komşu görüntülerin bulunması
2. Altyazıların bölütlenmesi
3. Oybırlığı Altyazısının belirlenmesi

# Komşu görüntülerin bulunması

- AlexNet (fc7) öznitelikleri
- Uyarlamalı komşu seçimi



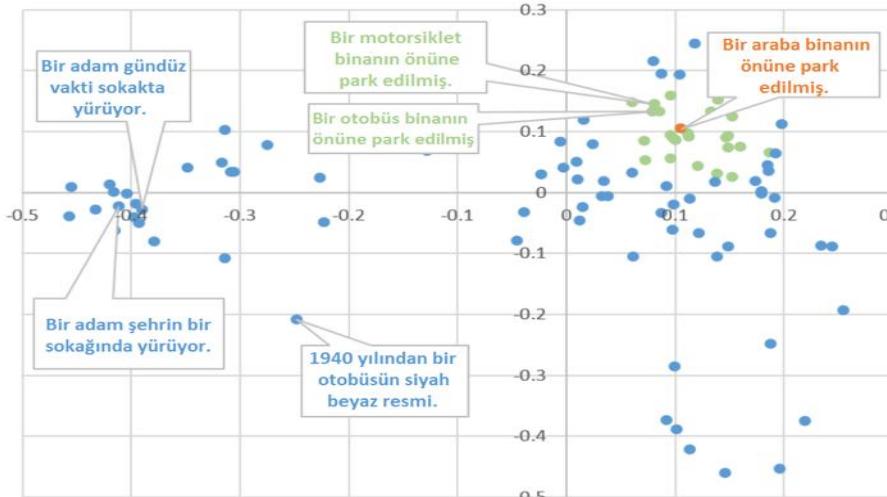
$$\begin{aligned} \mathcal{N}(I_q) &= \{(I_i, c_i) \mid \text{dist}(I_q, I_i) \leq (1 + \epsilon)\text{dist}(I_q, I_{\text{closest}}), \\ I_{\text{closest}} &= \arg \min \text{dist}(I_q, I_i), I_i \in \mathcal{T}\} \end{aligned} \quad (1)$$

# Altyazıların Bölütlenmesi

- Zemberek ile kelimelerin çekim eklerinin atılması

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

# Oybırlığı ile Altyazının Bulunması



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

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

- Oybırlığı altyazı:

$$c^* = \arg \max_{c \in C} \sum_{c' \in C} Sim(c, c')$$

- Benzerlik için BLEU metriği kullanılmıştır:

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



Kum pistte yarışan yarış köpekleri



Denizde koşan köpek



Gülümseyerek kameraya poz veren bir adam ve bir kadın



Frizbiyi havada yakalayan bir köpeği izleyen bir adam



Sarışın bir erkek çocuğu denize atlıyor.



Karda yanyana koşmakta olan iki süs köpeği

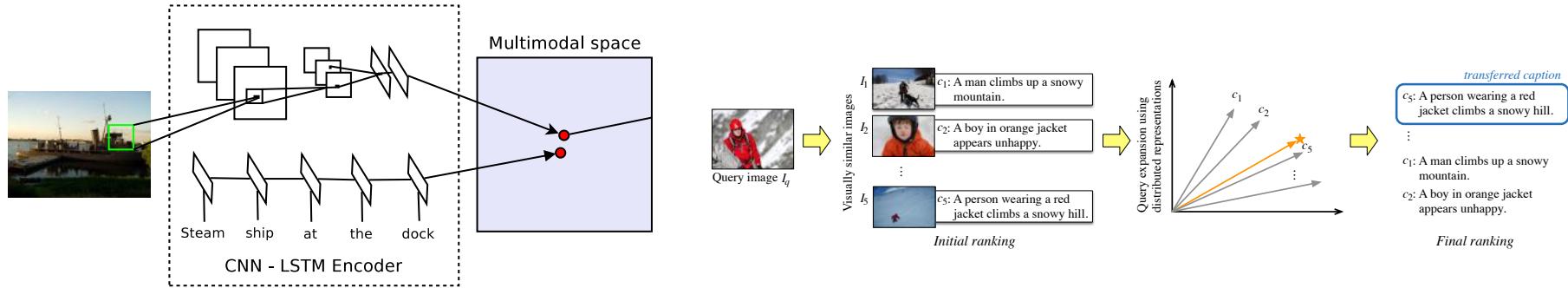


Endişeli endişeli bir yerlere bakmaka  
olan bir kadın



Bir futbol maçı

# Sıralama ve aktarma tabanlı görüntü altyazılıma



## • Üstün yönleri

- Daima dilbilgisine uygun açıklamalar üretmeleri,
- RNN veya LSTM tabanlı oluşturma yöntemlerine oranla gerçekleştiriminin görece basit olması

## • Zayıf yönleri

- Veri güdümlü (*data-driven*) yaklaşımalar oldukları için başarılarının çok kipli verinin büyüklüğü ile orantılı olması,
- Genelleme ve yeni durumlara (yeni görüntü içeriklerine) uyarlanabilirliklerinin görece zayıf oluşu

# Oluşturma Tabanlı Yaklaşımlar ve Diğer Uygulamalar

