

Hesaplamalı Anlambilim Ders Notları

Görsel Anlambilim

Mehmet Fatih AMASYALI
Hesaplamalı Anlambilim
Ders Notları
BÖLÜM 7



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

NOVA Research Lab

Hesaplamalı Anlambilim Ders Notları

İçerik

- Resim Tanımı Üretme
- Video Tanımlama
- Görsel Soru Cevaplama
- Karşıtsal Öğrenme



NOVA Research Lab

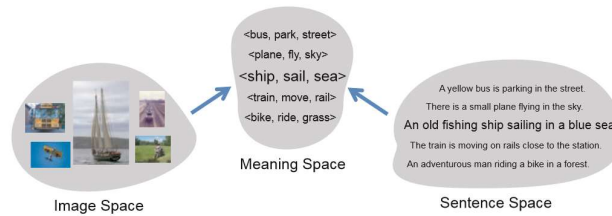
Resim Tanımı Üretme

- 2 yaklaşım
 - Resimden CNN vb.lerle objeleri, özelliklerini vb. bul. Bunlardan (dil modeli vb. ile) cümleleri üret.
 - Problemi diller arası çeviri olarak ele al (görsel bir dilden metne çeviri)



Every Picture Tells a Story *

- Birinci yaklaşıma örnek
- Resimleri ve cümleleri en benzediği (obje, eylem, yer) üçlüler ile ifade eder. Hazır listeler üzerinde çalışır. Resim ile hazır cümlelerden birini eşler.



Hesaplamalı Anlambilim Ders Notları

Every Picture Tells a Story

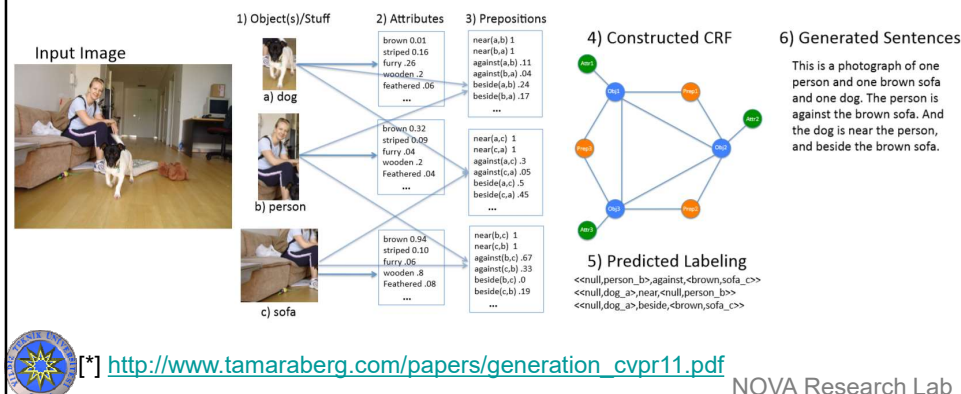
• (

(pet, sleep, ground) (dog, sleep, ground) (animal, sleep, ground) (animal, stand, ground) (goat, stand, ground)	see something unexpected. Cow in the grassfield. Beautiful scenery surrounds a fluffy sheep. Dog hearing sheep in open terrain. Cattle feeding at a trough.
(furniture, place, furniture) (furniture, place, room) (furniture, place, home) (bottle, place, table) (display, place, table)	Refrigerator almost empty. Foods and utensils. Eatables in the refrigerator. The inside of a refrigerator apples, cottage cheese, tupperwares and lunch bags. Squash apenny white store with a hand statue, picnic tables in front of the building.
(transportation, move, track) (bike, ride, track) (transportation, move, road) (pet, sleep, ground) (bike, ride, road)	A man stands next to a train on a cloudy day A backpacker stands beside a green train This is a picture of a man standing next to a green train There are two men standing on a rocky beach, smiling at the camera. This is a person laying down in the grass next to their bike in front of a strange white building.
(display, place, table) (furniture, place, furniture) (furniture, place, furniture) (bottle, place, table) (furniture, place, home)	This is a lot of technology. Somebody's screensaver of a pumpkin A black laptop is connected to a black Dell monitor This is a dual monitor setup Old school Computer monitor with way to many stickers on it

Hesaplamalı Anlambilim Ders Notları

BabyTalk *

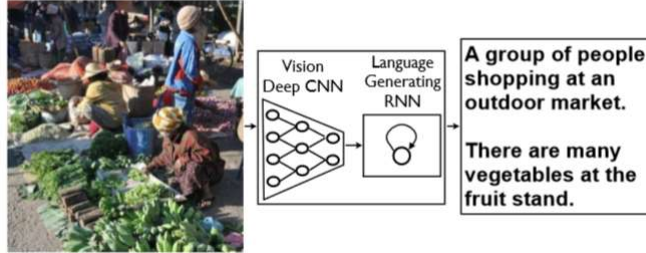
- Birinci yaklaşıma örnek
- (resim, tanım) ikililerine ihtiyaç duymaz. Hazır obje, özellik ve konum listeleri üzerinde çalışır.
- Deep learning öncesi, deformable part models + SVM



[*] http://www.tamaraberg.com/papers/generation_cvpr11.pdf

NOVA Research Lab

Show and Tell *



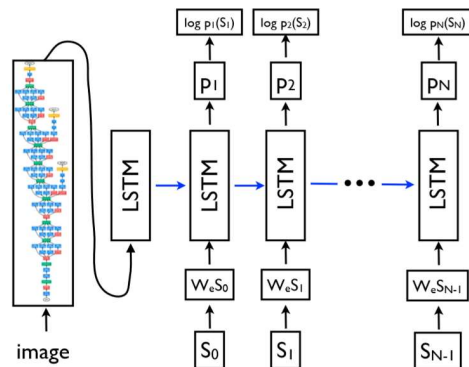
- 2. yaklaşım: Uçtan uca bir sistem
- Optimizasyon $\theta^* = \arg \max_{\theta} \sum_{(I,S)} \log p(S|I; \theta)$
- I: resim
- S: resmin doğru tanımı
- Θ : model parametreleri
- N uzunluğundaki bir tanımın olasılığı $\log p(S|I) = \sum_{t=0}^N \log p(S_t|I, S_0, \dots, S_{t-1})$

[*] <https://arxiv.org/pdf/1411.4555.pdf>

NOVA Research Lab

Resim Tanımı Üretme

- CNN çıkışı LSTM'e giriş
- CNN parametreleri pre-trained le başlıyor
- S_x : one hat vektör
- W_e : Word embedding layer
- 512 lstm hücresi var



NOVA Research Lab

Resim Tanımı Üretme

- Tanım üretimi:
 - Sampling: p1'e göre kelime seç, sonra p2, p3, ...
 - Beam search: O ana kadar ki en olası 20 cümleyi tut. Olasılık hesabı

$$S = \arg \max_{S'} p(S'|I)$$
 - Üretilen 20 cümleden en olası olanı seçildiğinde %80'i eğitim kümesindeki cümlelerden



Resim tanımları sonuçlar

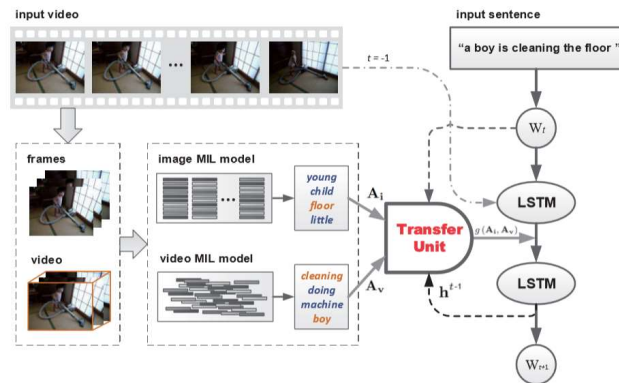


- Eğitim kümesine çok bağlı 😊



Video tanımlama *









- Yine çeviri sistemi tabanlı. 2D filtrelerle resimlerden, 3D filtrelerle video küplerinden (resimlerin arkaya doğru dizili hal) özellik çıkarıyor.



[*] http://openaccess.thecvf.com/content_cvpr_2017/papers/Pan_Video_Captioning_With_CVPR_2017_paper.pdf

NOVA Research Lab

Görsel Soru Cevaplama*

 <p>Is something under the sink broken?</p> <table border="1"> <tr><td>yes</td><td>no</td></tr> <tr><td>yes</td><td>no</td></tr> <tr><td>yes</td><td>no</td></tr> </table> <p>What number do you see?</p> <table border="1"> <tr><td>33</td><td>5</td></tr> <tr><td>33</td><td>6</td></tr> <tr><td>33</td><td>7</td></tr> </table>	yes	no	yes	no	yes	no	33	5	33	6	33	7	 <p>Can you park here?</p> <table border="1"> <tr><td>no</td><td>no</td><td>no</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> </table> <p>What color is the hydrant?</p> <table border="1"> <tr><td>white and orange</td><td>red</td></tr> <tr><td>white and orange</td><td>red</td></tr> <tr><td>white and orange</td><td>yellow</td></tr> </table>	no	no	no	no	no	yes	no	no	yes	white and orange	red	white and orange	red	white and orange	yellow	 <p>What kind of store is this?</p> <table border="1"> <tr><td>bakery</td><td>art supplies</td></tr> <tr><td>bakery</td><td>grocery</td></tr> <tr><td>pastry</td><td>grocery</td></tr> </table> <p>Is the display case as full as it could be?</p> <table border="1"> <tr><td>no</td><td>no</td><td>yes</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> </table>	bakery	art supplies	bakery	grocery	pastry	grocery	no	no	yes	no	no	yes	no	no	yes	 <p>How many bikes are there?</p> <table border="1"> <tr><td>2</td><td>3</td></tr> <tr><td>2</td><td>4</td></tr> <tr><td>2</td><td>12</td></tr> </table> <p>What number is the bus?</p> <table border="1"> <tr><td>48</td><td>4</td></tr> <tr><td>48</td><td>46</td></tr> <tr><td>48</td><td>number 6</td></tr> </table>	2	3	2	4	2	12	48	4	48	46	48	number 6									
yes	no																																																																	
yes	no																																																																	
yes	no																																																																	
33	5																																																																	
33	6																																																																	
33	7																																																																	
no	no	no																																																																
no	no	yes																																																																
no	no	yes																																																																
white and orange	red																																																																	
white and orange	red																																																																	
white and orange	yellow																																																																	
bakery	art supplies																																																																	
bakery	grocery																																																																	
pastry	grocery																																																																	
no	no	yes																																																																
no	no	yes																																																																
no	no	yes																																																																
2	3																																																																	
2	4																																																																	
2	12																																																																	
48	4																																																																	
48	46																																																																	
48	number 6																																																																	
 <p>Does this man have children?</p> <table border="1"> <tr><td>yes</td><td>yes</td><td>yes</td></tr> <tr><td>yes</td><td>yes</td><td>yes</td></tr> <tr><td>yes</td><td>yes</td><td>yes</td></tr> </table> <p>Is this man crying?</p> <table border="1"> <tr><td>no</td><td>no</td><td>yes</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> <tr><td>no</td><td>no</td><td>yes</td></tr> </table>	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	yes	no	no	yes	no	no	yes	 <p>Has the pizza been baked?</p> <table border="1"> <tr><td>yes</td><td>yes</td><td>yes</td></tr> <tr><td>yes</td><td>yes</td><td>yes</td></tr> <tr><td>yes</td><td>yes</td><td>yes</td></tr> </table> <p>What kind of cheese is topped on this pizza?</p> <table border="1"> <tr><td>feta</td><td>mozzarella</td></tr> <tr><td>ricotta</td><td>mozzarella</td></tr> <tr><td>ricotta</td><td>mozzarella</td></tr> </table>	yes	yes	yes	yes	yes	yes	yes	yes	yes	feta	mozzarella	ricotta	mozzarella	ricotta	mozzarella	 <p>How many pickles are on the plate?</p> <table border="1"> <tr><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td></tr> </table> <p>What is the shape of the plate?</p> <table border="1"> <tr><td>circle</td><td>circle</td></tr> <tr><td>round</td><td>round</td></tr> <tr><td>round</td><td>round</td></tr> </table>	1	1	1	1	1	1	circle	circle	round	round	round	round	 <p>What does the sign say?</p> <table border="1"> <tr><td>stop</td><td>stop</td><td>stop</td></tr> <tr><td>stop</td><td>stop</td><td>yield</td></tr> <tr><td>stop</td><td>stop</td><td>yield</td></tr> </table> <p>What shape is this sign?</p> <table border="1"> <tr><td>octagon</td><td>octagon</td><td>diamond</td></tr> <tr><td>octagon</td><td>octagon</td><td>octagon</td></tr> <tr><td>octagon</td><td>octagon</td><td>round</td></tr> </table>	stop	stop	stop	stop	stop	yield	stop	stop	yield	octagon	octagon	diamond	octagon	octagon	octagon	octagon	octagon	round
yes	yes	yes																																																																
yes	yes	yes																																																																
yes	yes	yes																																																																
no	no	yes																																																																
no	no	yes																																																																
no	no	yes																																																																
yes	yes	yes																																																																
yes	yes	yes																																																																
yes	yes	yes																																																																
feta	mozzarella																																																																	
ricotta	mozzarella																																																																	
ricotta	mozzarella																																																																	
1	1																																																																	
1	1																																																																	
1	1																																																																	
circle	circle																																																																	
round	round																																																																	
round	round																																																																	
stop	stop	stop																																																																
stop	stop	yield																																																																
stop	stop	yield																																																																
octagon	octagon	diamond																																																																
octagon	octagon	octagon																																																																
octagon	octagon	round																																																																

Yeşiller resmi görerek, maviler görmeden verilen cevaplar



[*] <http://visualqa.org/>

NOVA Research Lab

Gerekli kabiliyetler *

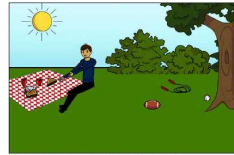
- Nesne tanıma:
 - What kind of cheese is on the pizza?
 - How many bikes are there?
- Aktivite tanıma
 - Is this man crying?
- Hayat bilgisi ve Çıkarım
 - Is this a vegetarian pizza?
 - Does this person have 20/20 vision?
 - Is this person expecting company?



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?



[*] <http://arxiv.org/pdf/1505.00468v6.pdf>

NOVA Research Lab

Yöntem

- Frekansı en yüksek olan ilk (K=1000) cevap (1000 sınıflı bir problem)
- Sorular bag of words (1030), resimler VGGnet (4096) ile kodlanmış
- MLP ve LSTM kullanılmış
- Giriş: soru, resim, soru + resim
- Baseline sonuçları: (resimlerin etkisi çok değil !)

	Open-Answer				Multiple-Choice			
	All	Yes/No	Number	Other	All	Yes/No	Number	Other
Question	48.09	75.66	36.70	27.14	53.68	75.71	37.05	38.64
Image	28.13	64.01	00.42	03.77	30.53	69.87	00.45	03.76
Q+I	52.64	75.55	33.67	37.37	58.97	75.59	34.35	50.33
LSTM Q	48.76	78.20	35.68	26.59	54.75	78.22	36.82	38.78
LSTM Q+I	53.74	78.94	35.24	36.42	57.17	78.95	35.80	43.41



Yarış sürüyor <http://visualqa.org/>

NOVA Research Lab

VQA veri kümesinin oluşturulması

- Sorular üretilirken dikkatli bir yol izlenmiş:
- Bir resim verilir, bir robotun zor cevaplayacağı sorular üretilmesi istenmiş
- *We have built a smart robot. It understands a lot about images. It can recognize and name all the objects, it knows where the objects are, it can recognize the scene (e.g., kitchen, beach), people's expressions and poses, and properties of objects (e.g., color of objects, their texture). Your task is to stump this smart robot!*



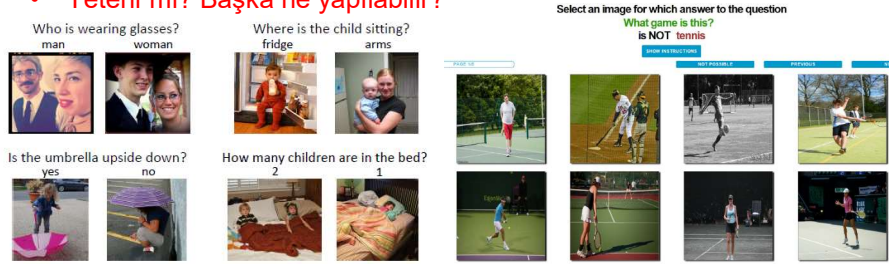
Eleştiriler

- **Ancak** VQA veri kümesinde
- «What sport is» ile başlayan soruların %41'inin cevabı «tenis»
- «How many» ile başlayan soruların %39'unun cevabı «2»
- «Bu resimde balık var mı?» sorusu ancak resimde balık varsa soruluyor. «bu resimde X var mı?» sorularının %81'inin cevabı «evet»



Veri kümesinde güncelleme*

- Sorularla cevaplar arasındaki düzenlilikleri ortadan kaldırmak için
- Resme bakmadan doğru cevap verebilmeyi önlemek için
- Bu tarz sistemlerin başarılarının abartılmaması için
- Aynı sorunun farklı resimler için farklı cevaplarının olduğu bir benchmark
- **Yeterli mi? Başka ne yapılabilir?**



[*] <https://arxiv.org/pdf/1612.00837.pdf>

NOVA Research Lab

Karşıtsal Öğrenme Contrastive Learning

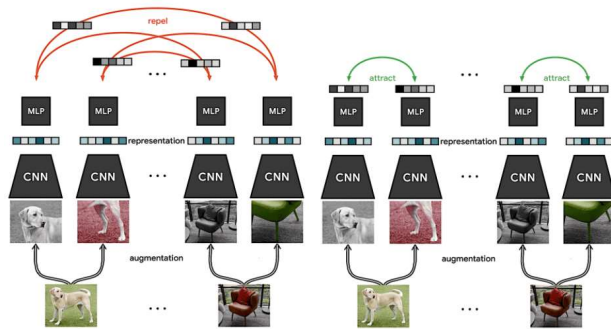
- 2 girdi için benzer ya da değil diyebilmek
- Çözüm: benzer olanlar için benzer, farklı olanlar için farklı çıktıları üretebilmek
- Eğitim kümesini oluşturmada eğitici ve eğitici olmayan yöntemler var.
- Eğitici: Farklı 2 kedi resmi benzesin, 1 kedi ve 1 balık benzemesin.



NOVA Research Lab

simCLR*

Self supervised, eğiticiyiz ☺, eğitim kümesi üretme yöntemi ölçeklenebilir ☺
Aynı resim içindeki resim parçaları (transform edilmiş halleri) benzer temsillere,
farklı resimlerdeki farklı temsillere sahip olsun
Eğitim sonrası temsiller istenilen görevde kullanılabilir.



Sorular:

Resim-metin ikilileri
üzerinde
nasıl çalıştırabiliriz?

Bir cevap:

<https://openai.com/blog/clip/>
Peki sadece metinler
üzerinde?



[*] <https://arxiv.org/abs/2002.05709>

[**] <https://amitness.com/2020/03/illustrated-simclr/>

[***] https://docs.google.com/presentation/d/1ccddJFD_j3p3h0TCqSV9ajSi2y1yOfh0-LJoK29ircs

[****] <https://www.eye-on.ai/s/Geoff-Hinton-final-transcript.docx>

NOVA Research Lab

Yorum

- 5 yıl içinde büyük gelişim
- Sebepleri
 - Nesne tanıyıcıların başarısının artışı
 - Etiketli veri kümelerinin artışı
 - Self supervised yöntemlerdeki gelişmeler
- Dikkat: başarıları / veri kümelerini iyi incelemek gerekiyor
- Aşağıda neler oluyor ☺



NOVA Research Lab