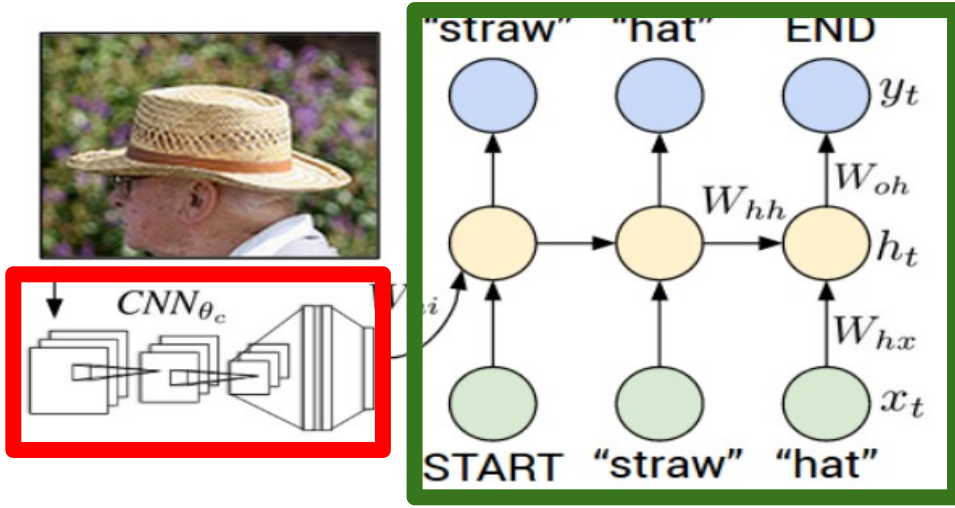
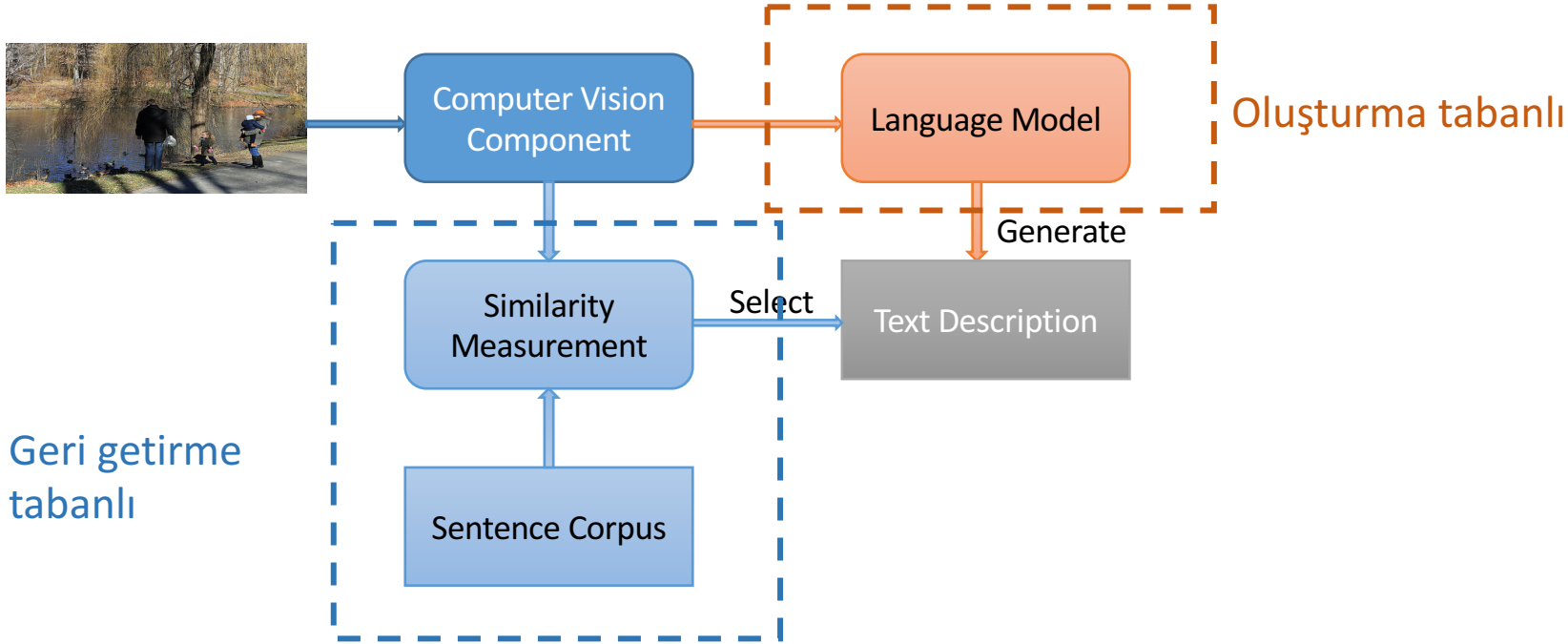
Oluşturma Tabanlı Yaklaşımlar ve Diğer Uygulamalar

Görüntü altyazılama

- Sıralama tabanlı yaklaşımlar
- Aktarma tabanlı yaklaşımlar
- **Oluşturma tabanlı yaklaşımlar**



Genel Yaklaşım (Derin Öğrenmesiz)



Yansı: Hongzhi Li

Baby Talk: Understanding and Generating Simple Image Descriptions

Kulkarni, et al., CVPR 2011

“This picture shows one person,

Kulkarni et al, CVPR11

"This picture shows one person, one grass,

Kulkarni et al, CVPR11

"This picture shows one person, one grass, one chair,

Kulkarni et al, CVPR11

“This picture shows one person, one grass, one chair, and one potted plant.

Kulkarni et al, CVPR11

“This picture shows one person, one grass. one chair. and one potted plant. The person is near the green grass,

Kulkarni et al, CVPR11

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair.

Kulkarni et al, CVPR11

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair,

Kulkarni et al, CVPR11

"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

Kulkarni et al, CVPR11

Algoritmanın İnsan Tanımlamalarıyla Karşılaştırılması

Algoritma:

"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

İnsan:

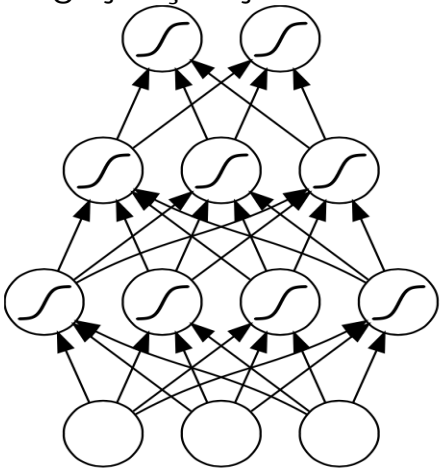
H1: A Lemonaide stand is manned by a blonde child with a cookie.
H2: A small child at a lemonade and cookie stand on a city corner.
H3: Young child behind lemonade stand eating a cookie.

Kulağa doğal gelmiyor!

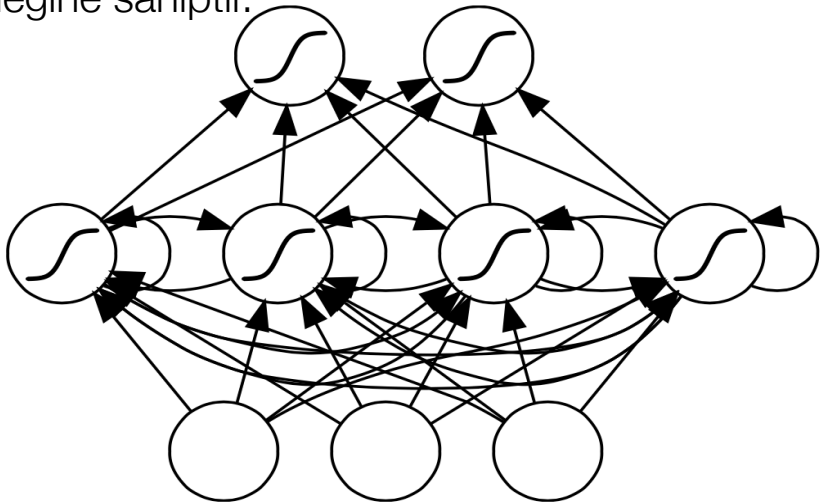
Kulkarni et al, CVPR11

Yinelemeli Sinir Ağları (Recurrent Neural Networks - RNN)

- Çok katmanlı perseptronlar sadece girdi-çıktı arasında modelleme yaparken, RNN yapıları bütün bir girdi geçmişini çıktı üzerinde modelleme yeteneğine sahiptir.



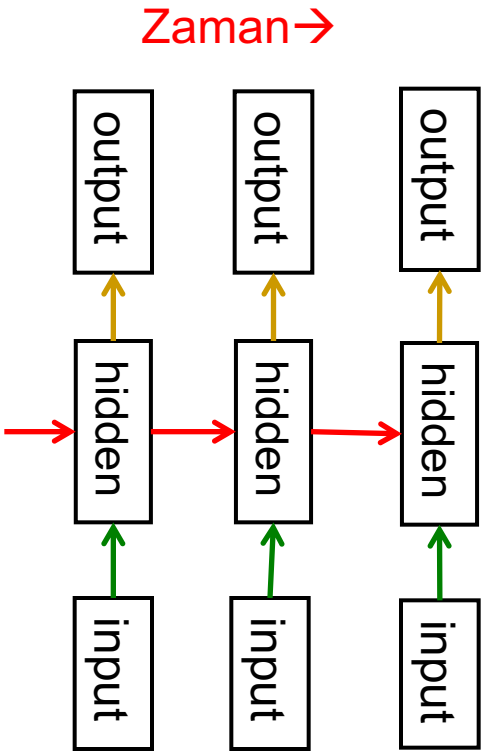
Çok katmanlı perseptron



Yinelemeli Ağ

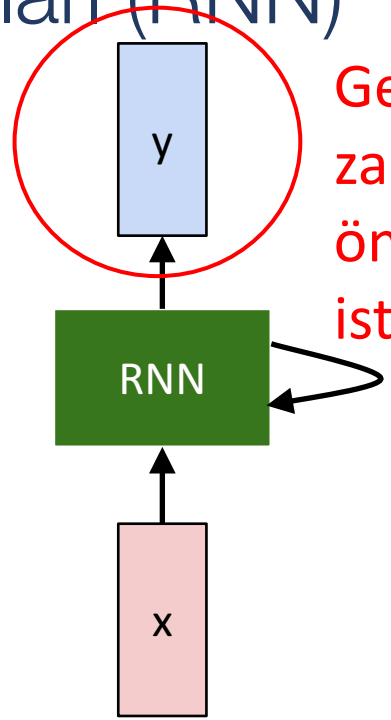
Yinelemeli Sinir Ağları

- Yinelemeli sinir ağları, iki temel özelliği sayesinde çok güçlü modeller oluşturma yeteneğine sahiptir.
  - Dağıtık saklı durum: geçmiş hakkında çok fazla bilginin verimli bir şekilde tutulmasına olanak sağlar.
  - Doğrusal olmayan dinamikler: saklı durumların kompleks şekilde güncellenmesini sağlar.



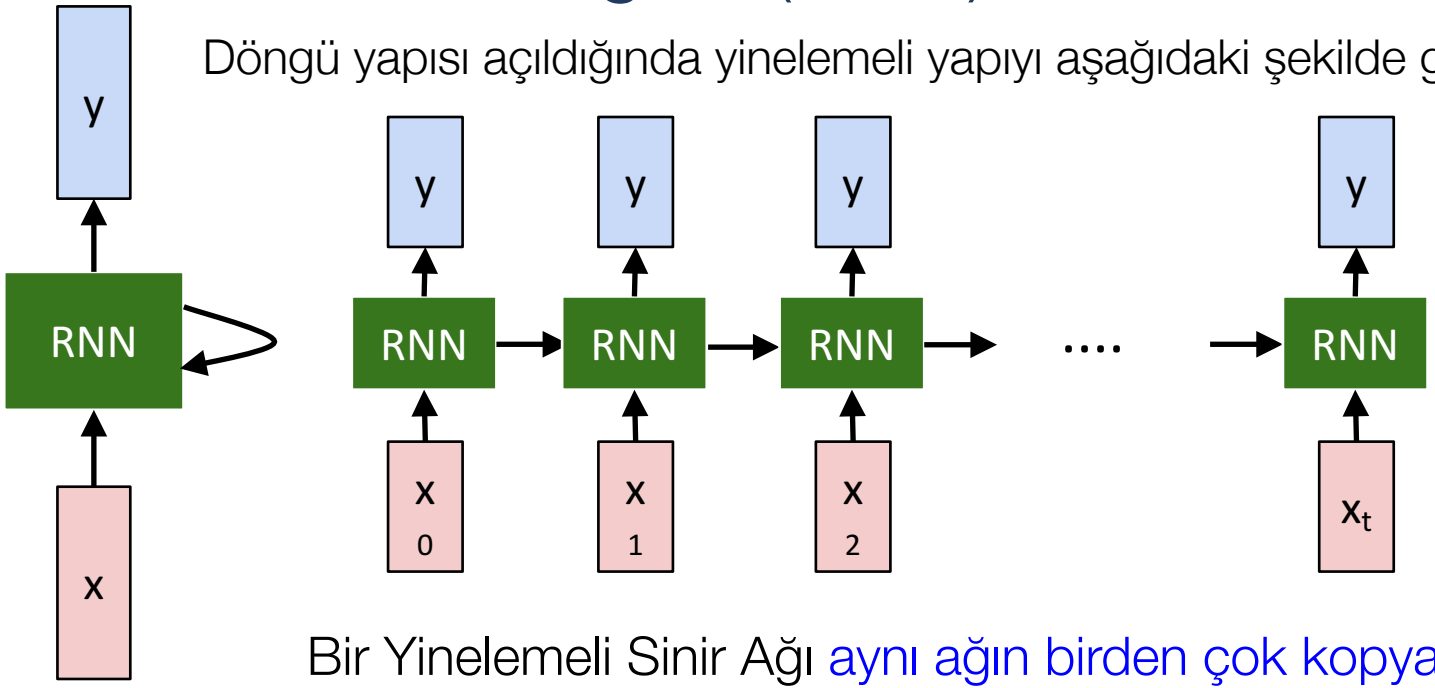
Yansı: G. Hinton

Yinelemeli Sinir Ağları (RNN)



Genellikle bazı zaman dilimlerinde öngörü yapılmak istenmektedir.

Yinelemeli Sinir Ağları (RNN)



Döngü yapısı açıldığında yinelemeli yapıyı aşağıdaki şekilde gözlemleyebiliriz:

Bir Yinelemeli Sinir Ağı aynı ağın birden çok kopyası olarak düşünülebilir, her bir alt ağ, ardışık ağa mesaj iletmektedir.

Yinelemeli Sinir Ağları (RNN)

Bir dizi x vektörü, her zaman diliminde yineleme formülü uygulanarak işlenir:

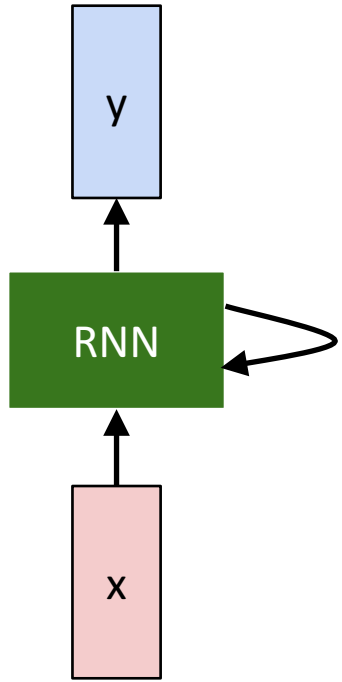
$$h_t = f_W(h_{t-1}, x_t)$$

Yeni durum

eski durum

t zamanındaki girdi vektörü

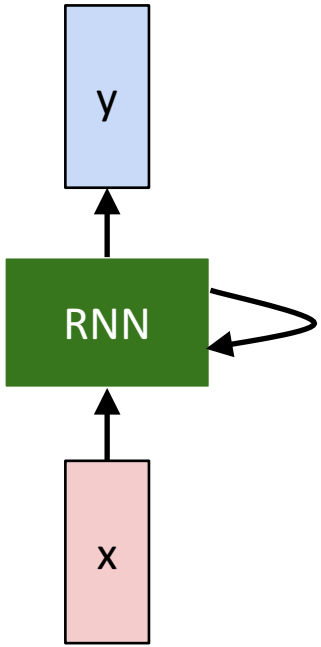
W parametrelerine sahip bir fonksiyon



Önemli: her zaman diliminde, aynı fonksiyon ve aynı parametreler kullanılmaktadır.

(En Temel) Yinelemeli Sinir Ağı

Durum tek bir *"saklı"* h vektöründen oluşmaktadır:



$$h_t = f_W(h_{t-1}, x_t)$$

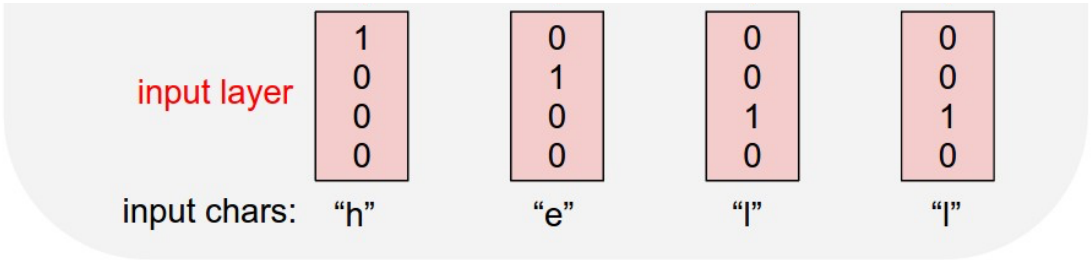
$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

Karakter seviyesindeki dil modeli örneği

Dağarcık:
[h,e,l,o]

Örnek eğitim verisi:
“hello”



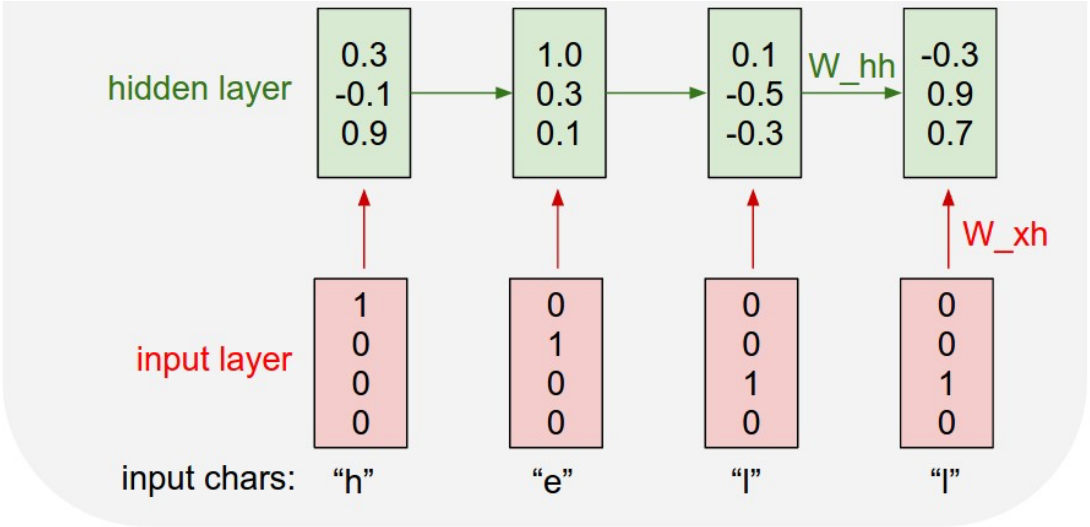
Yansı: Johnson, Karpathy, Fei Fei

Karakter seviyesindeki dil modeli örneği

Dağarcık:
[h,e,l,o]

Örnek eğitim verisi:
“hello”

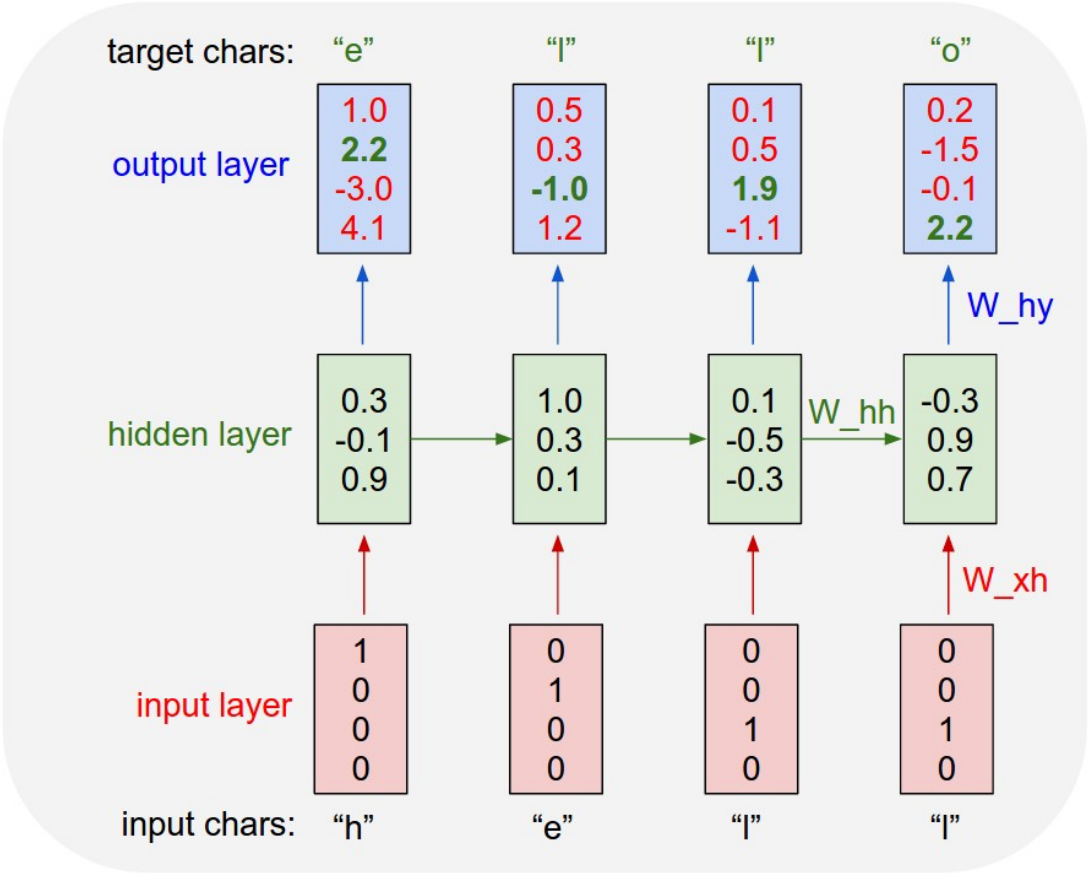
$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$



Karakter seviyesindeki dil modeli örneği

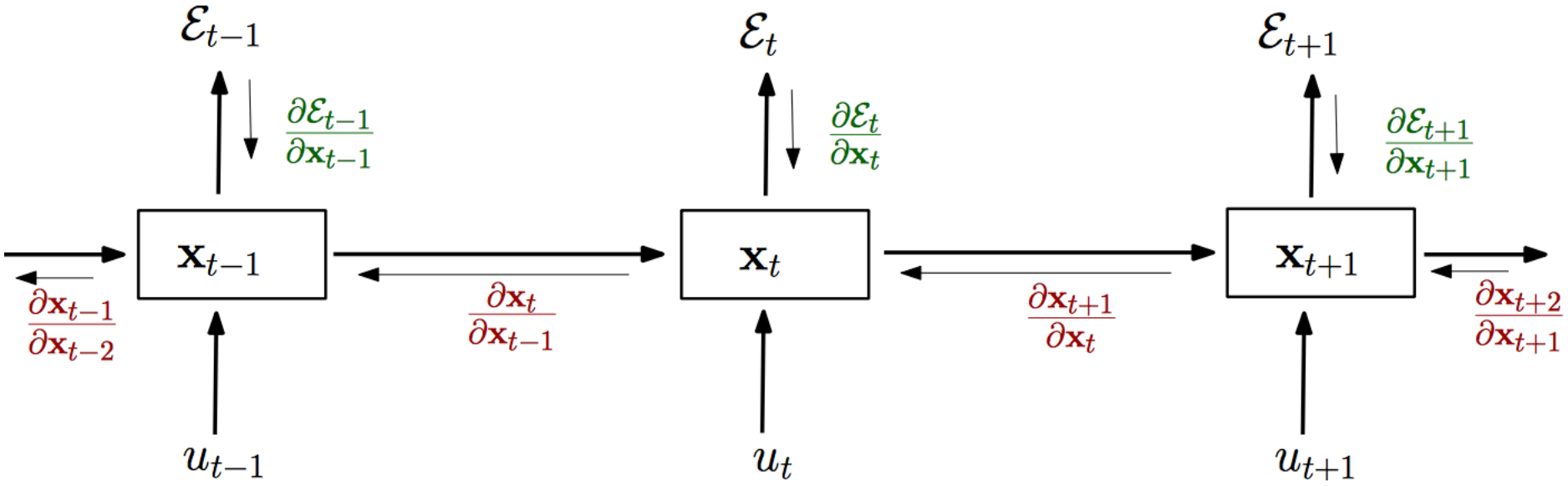
Dağarcık:
[h,e,l,o]

Örnek eğitim verisi:
“hello”

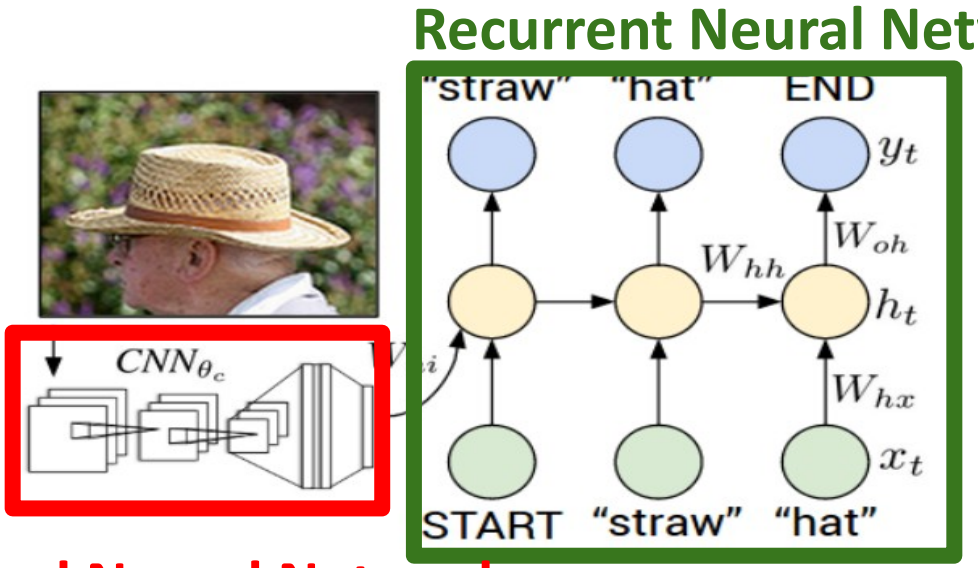


Zamanda geri yayılma (BPTT)

- Yinelemeli sinir ağı çoklu katman yapısında ifade edilir ve bu açılımlı modelde geri yayılım zaman ekseninde uygulanır.



RNN ile imge altyazılama



- Explain Images with Multimodal Recurrent Neural Networks, Mao et al., NIPS 2014.
- Deep Visual-Semantic Alignments for Generating Image Descriptions, Karpathy and Fei-Fei, CVPR 2015.
- Show and Tell: A Neural Image Caption Generator, Vinyals et al., CVPR 2015.
- Long-term Recurrent Convolutional Networks for Visual Recognition and Description, Donahue et al., CVPR 2015.
- Learning a Recurrent Visual Representation for Image Caption Generation, Chen and Zitnick, CVPR 2015.

Yansı: Andrej Karpathy

test image

Yansı: Andrej Karpathy

image

conv-64
conv-64
maxpool

conv-128
conv-128
maxpool

conv-256
conv-256
maxpool

conv-512
conv-512
maxpool

conv-512
conv-512
maxpool

FC-4096
FC-4096
FC-1000
softmax

test image

Yansı: Andrej Karpathy

image

conv-64

conv-64

maxpool

conv-128

conv-128

maxpool

conv-256

conv-256

maxpool

conv-512

conv-512

maxpool

conv-512

conv-512

maxpool

FC-4096

FC-4096

FC-1000

softmax

X

test image

Yansı: Andrej Karpathy

image

conv-64
conv-64
maxpool

conv-128
conv-128
maxpool

conv-256
conv-256
maxpool

conv-512
conv-512
maxpool

conv-512
conv-512
maxpool

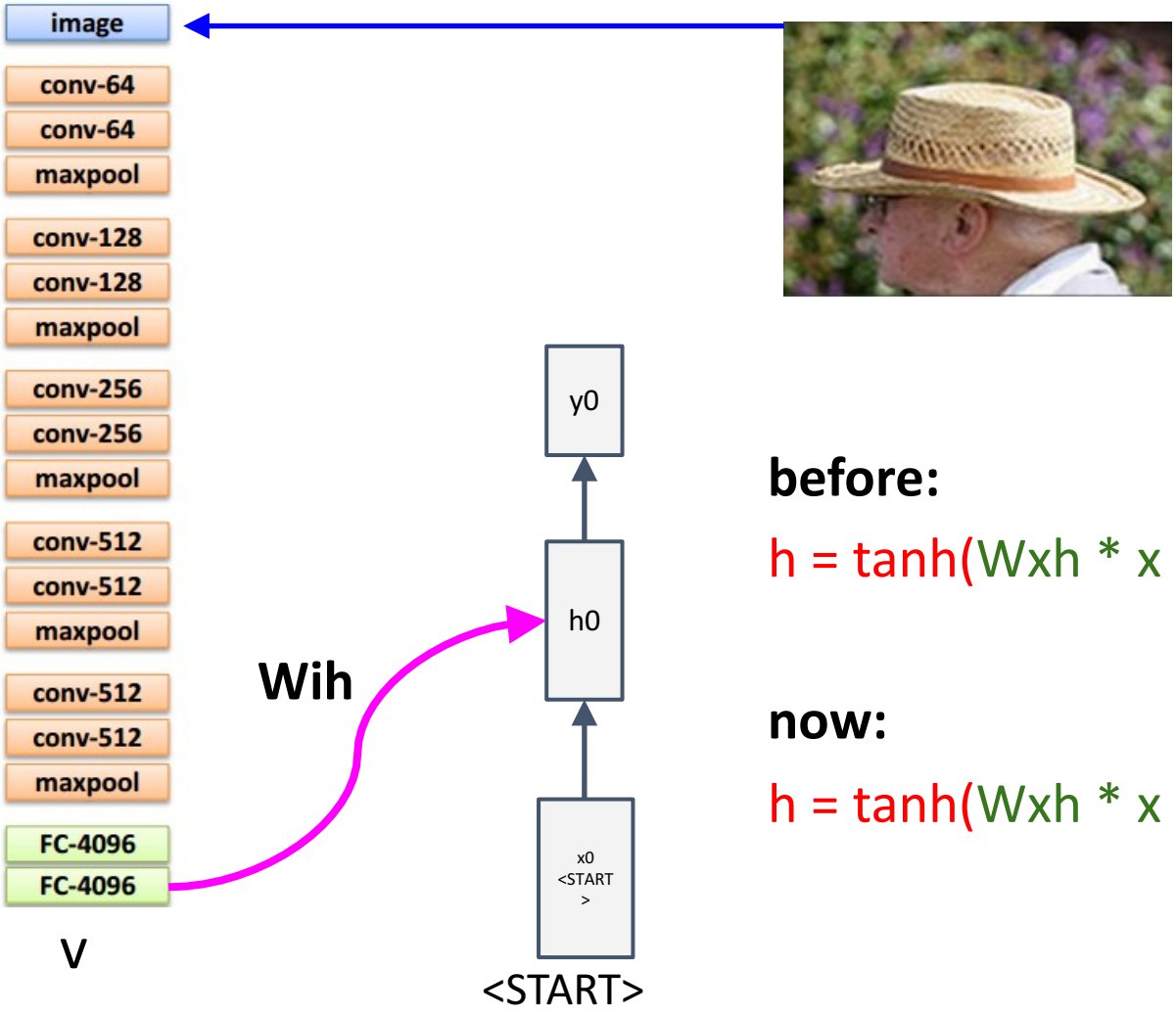
FC-4096
FC-4096

test image

x0
<START
>

<START>

Yansı: Andrej Karpathy



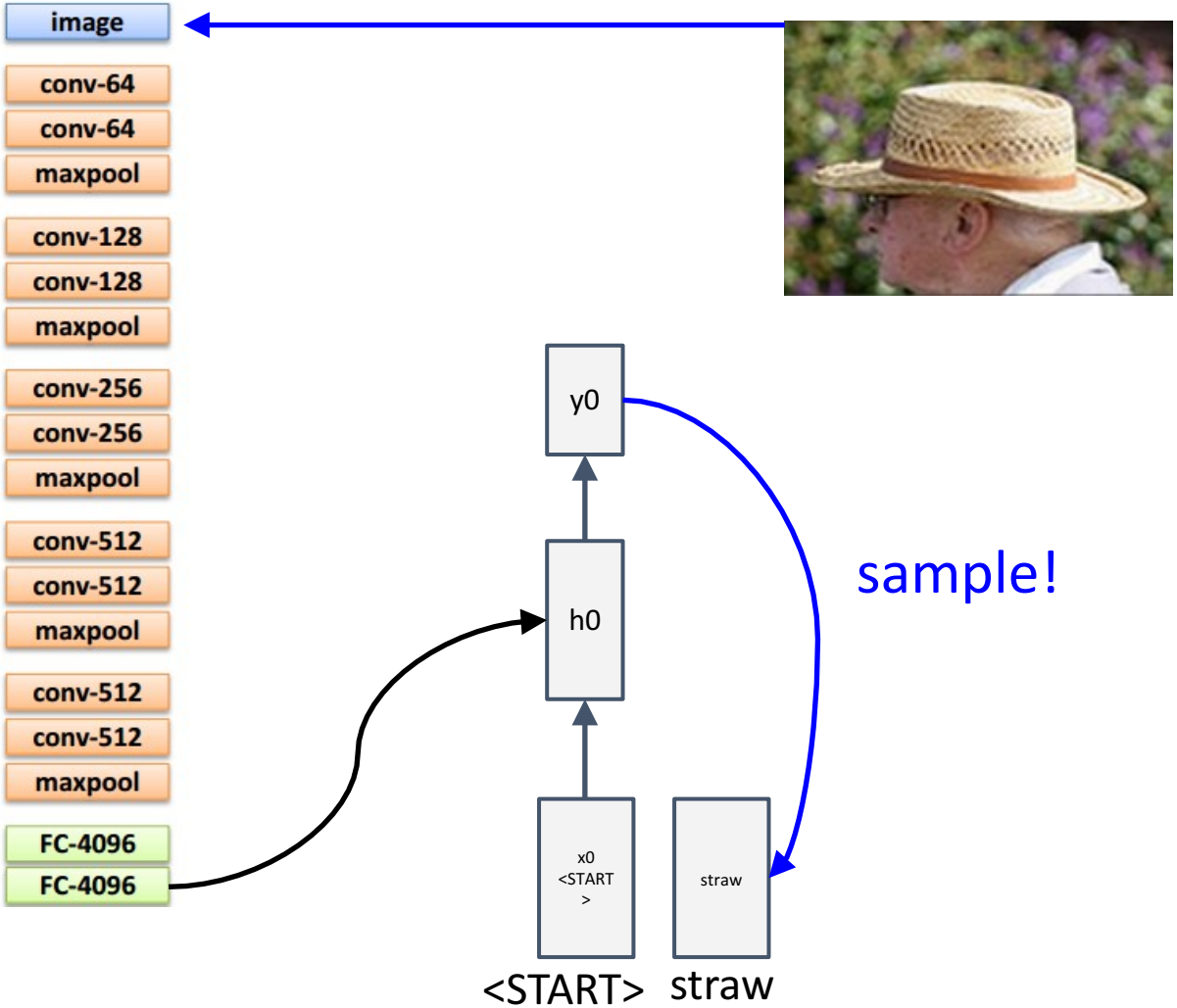
before:

$h = \tanh(W_{xh} * x + W_{hh} * h)$

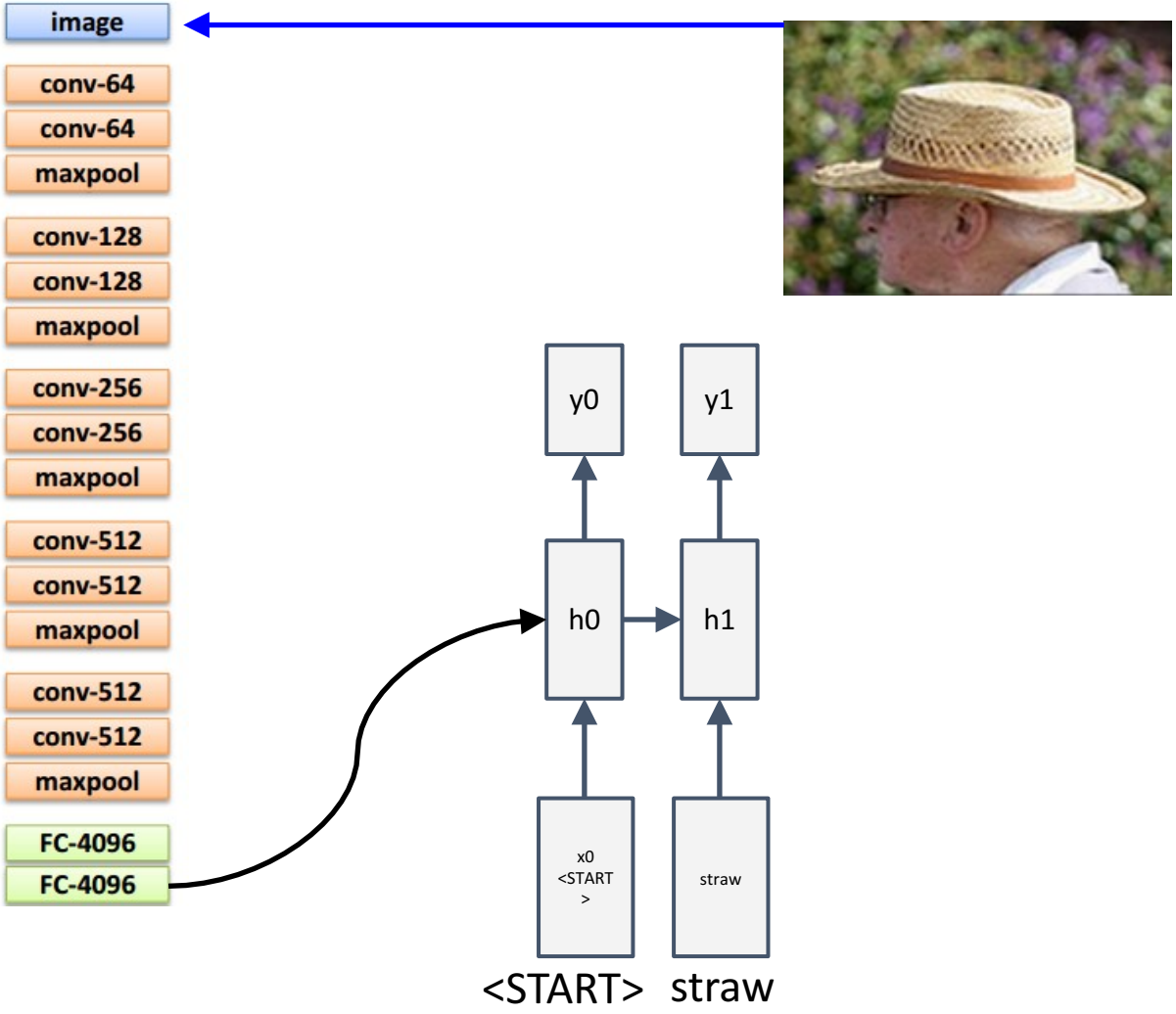
now:

$h = \tanh(W_{xh} * x + W_{hh} * h + \mathbf{W_{ih}} * \mathbf{v})$

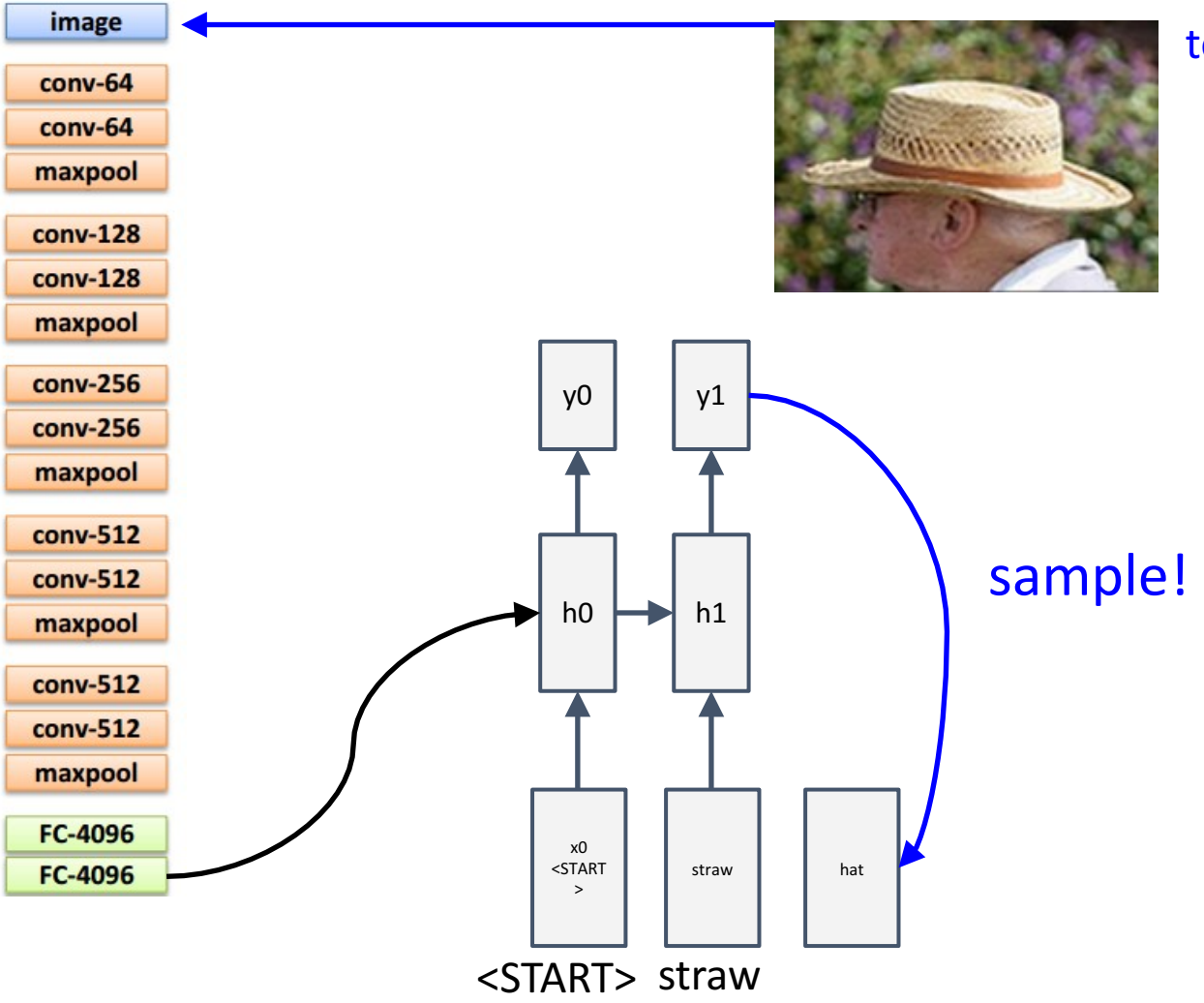
Yansı: Andrej Karpathy



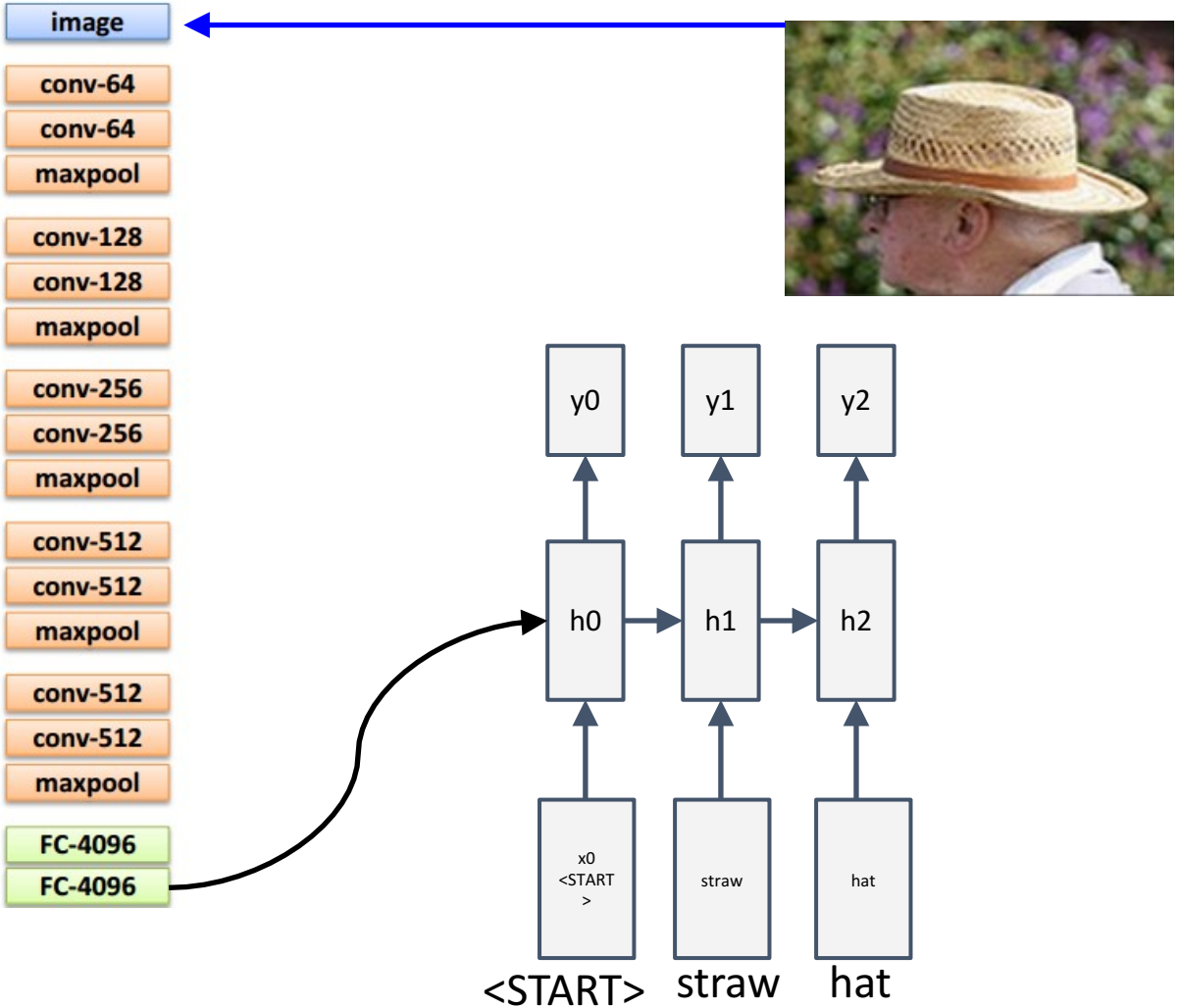
Yansı: Andrej Karpathy



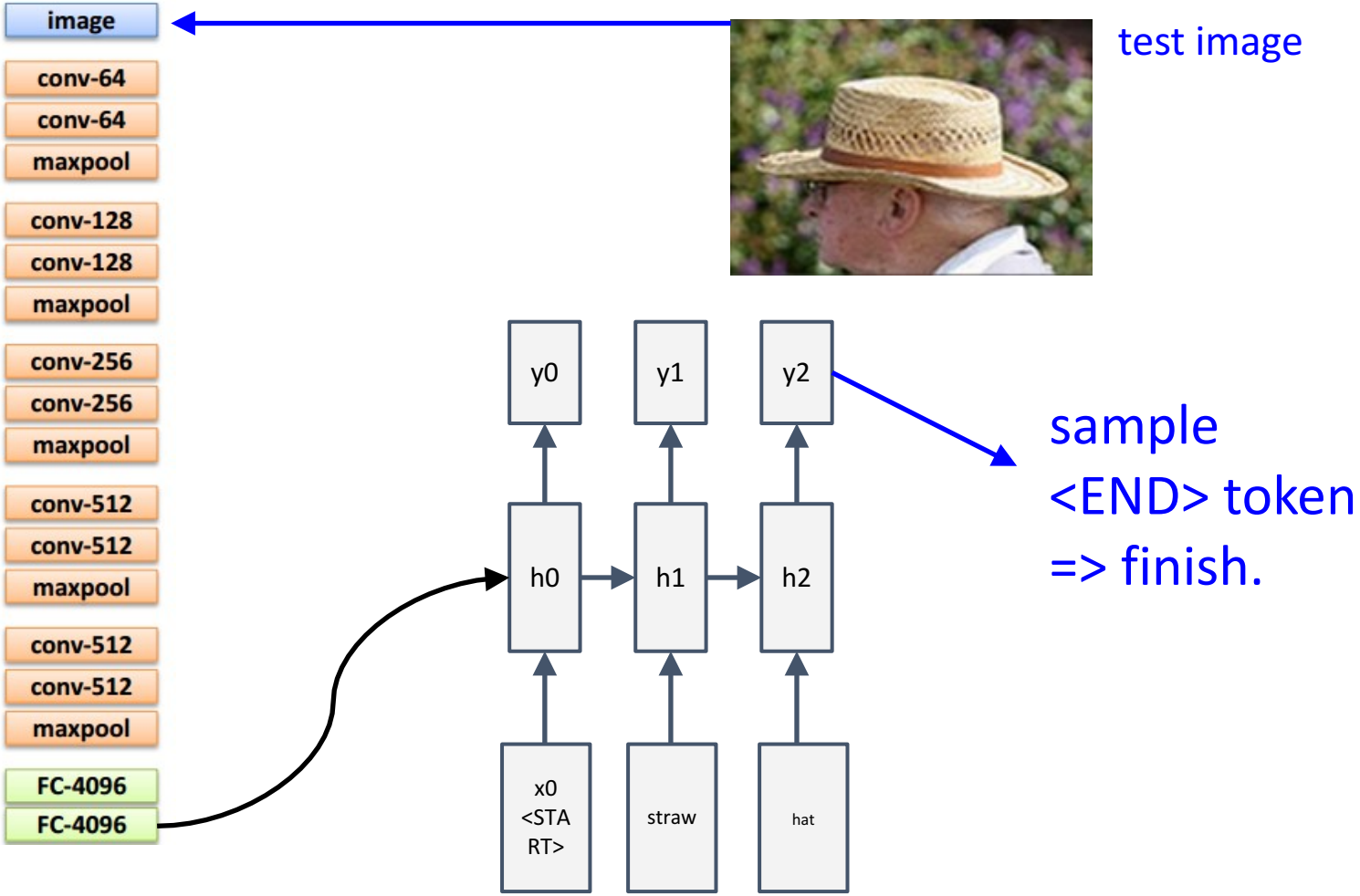
Yansı: Andrej Karpathy



Yansı: Andrej Karpathy



Yansı: Andrej Karpathy



Yansı: Andrej Karpathy

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

Yansı: Andrej Karpathy

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"a young boy is holding a baseball bat."

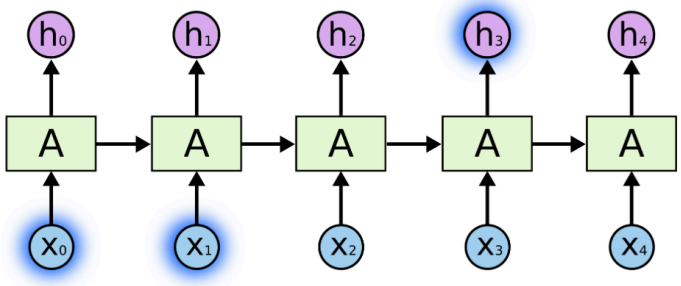
"a cat is sitting on a couch with a remote control."

"a woman holding a teddy bear in front of a mirror."

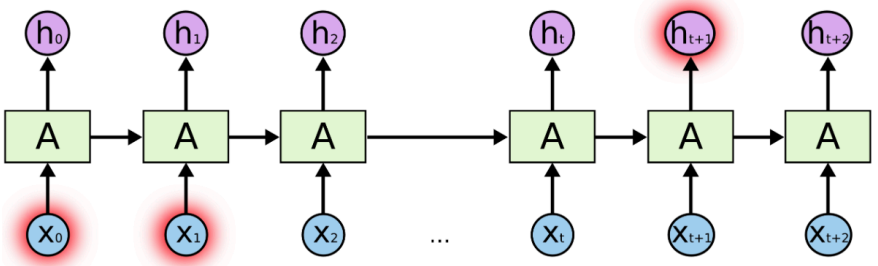
"a horse is standing in the middle of a road."

Uzun süreli bağımlılıklar problemi

- Temel RNN'ler önceki adımdaki bilgiyi şu anki bilgiye bağlarlar:
- - bu bir sonraki kelimeyi bulmak için yeterlidir: "gökyüzü bugün çok bulutlu"

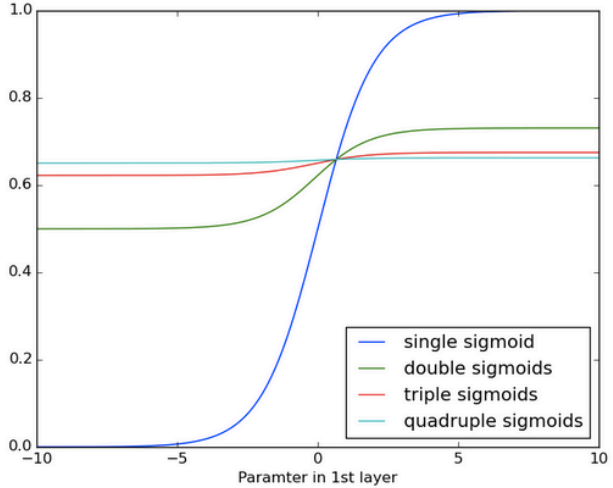


- - ama daha fazla bağlam bilgisi gerektiğinde yeterli değildir:
- "Fransada büyüdüm... Bu nedenle çok iyi konuşurum *Fransızcayı*."

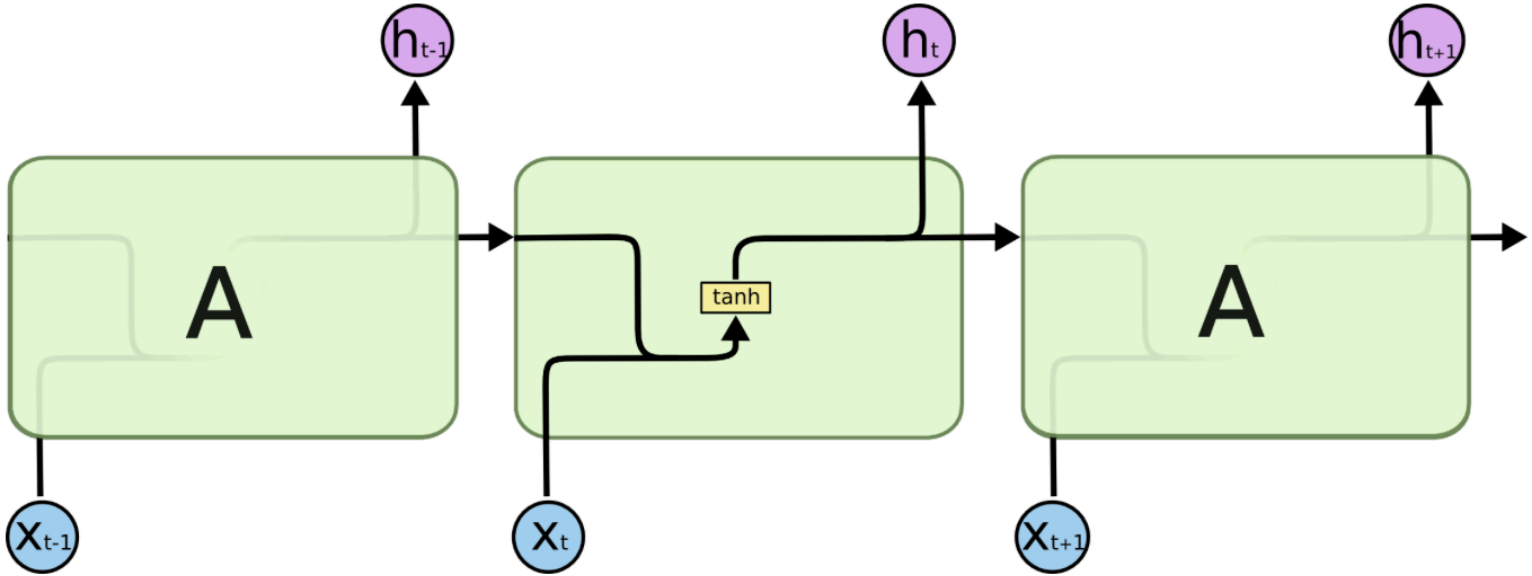


Kaybolan gradyan problemi

- **Geleneksel yinelemeli sinir ağlarında, gradyan geri yayılımı evresinde, gradyan sinyali çok kere çarpıma uğrayabilir**
- **Gradyanlar büyük ise**
  - **Gradyanlar daha da büyür ve öğrenme sapar**
  - **Çözüm: Gradyanları sabit bir max değerde kes**
- **Gradyanlar küçük ise**
  - **Kaybolan gradyanlar, öğrenme çok yavaşlar ya da durur**
  - **Çözüm: LSTM, GRU, ile bellek tanımlama**



Bütün yinelemeli sinir ağları tekrarlayan nöron zincirinden oluşmaktadır. Geleneksel RNN'de bu tekrarlayan yapıtaşı, girdi ve durum vektörlerinin doğrusal olmayan bileşimi şeklindedir.

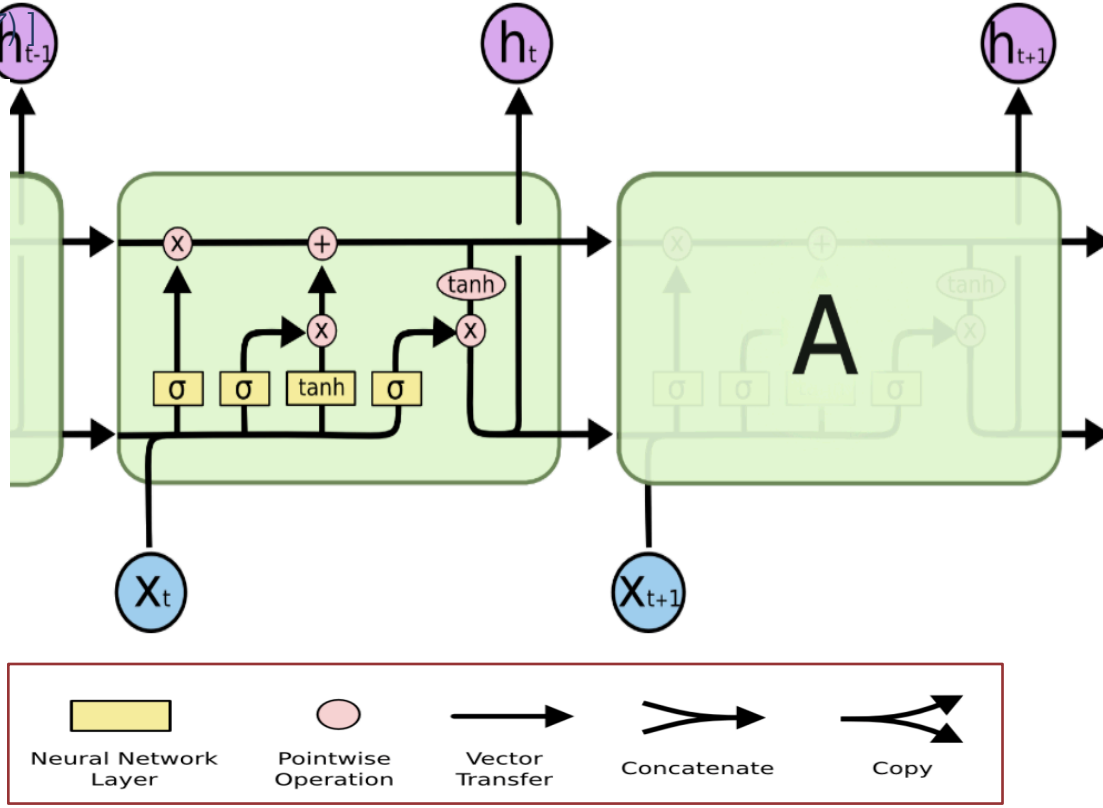


The repeating module in a standard RNN contains a single layer.

Kaynak: C. Olah

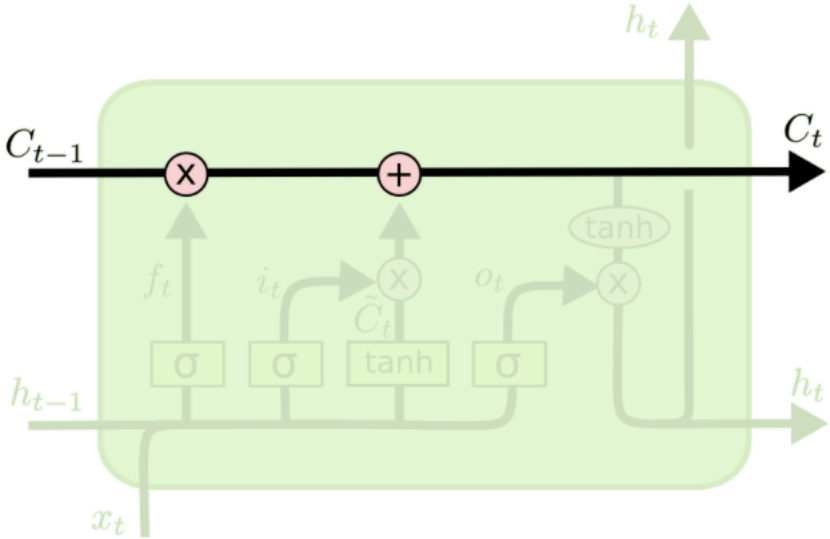
Uzun Kısa Süreli Bellek (Long Short Term Memory (LSTM))[Hochreiter & Schmidhuber (1997)]

- Lojistik ve lineer ünitelerin çarpımsal etkileşimleri ile oluşturulmuş bir bellek hücresidir:
- "input" kapısı açık durumda ise bilgi bellek hücresine girer.
- "forget" kapısı açık olduğu sürece bilgi, bellek hücresinde kalır.
- "output" kapısının açılması sureti ile bilgi hücre yapısından sonraki hücrelere aktarılabilir.

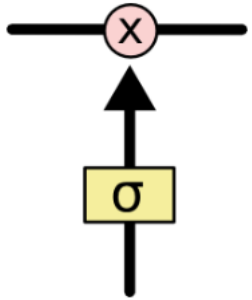


Kaynak: G Hinton and C. Olah

LSTM arkasındaki temel fikir : Hücre durumu



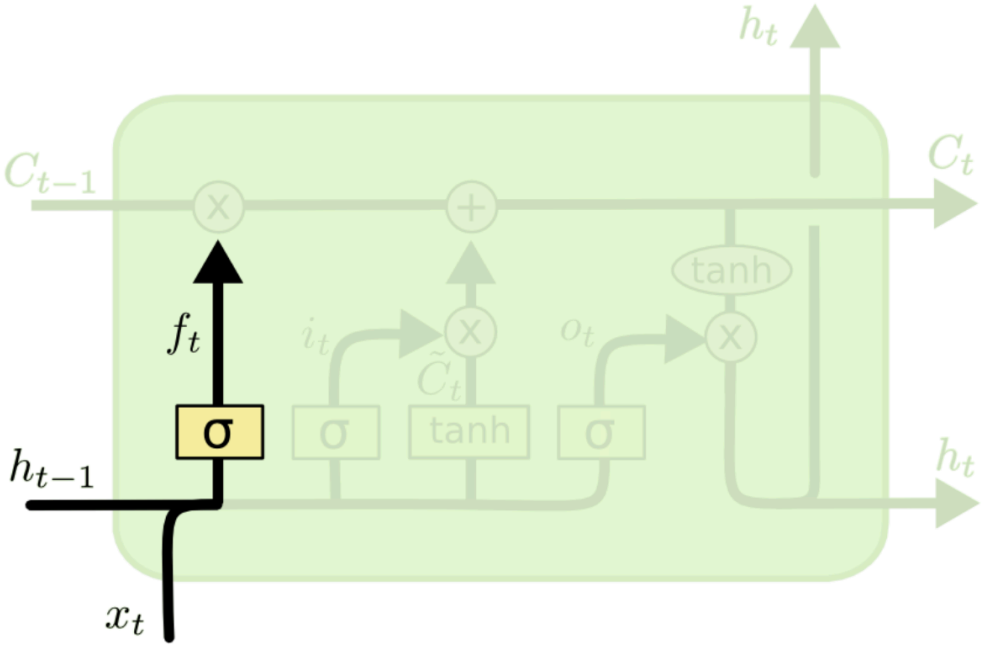
Kapılar, bilgiyi seçici olarak kullanmamıza yarayan yapılardır. Kapı fonksiyonları, tek bir sigmoid katmanından ve noktasal çarpım işleminden oluşur.



Bir LSTM'de bu kapılardan üç tane bulunmaktadır, ve bu kapılar hücre durumunu korumak ve kontrol etmek için kullanılır.

Kaynak: C. Olah

LSTM : Unutma kapısı (Forget Gate)



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

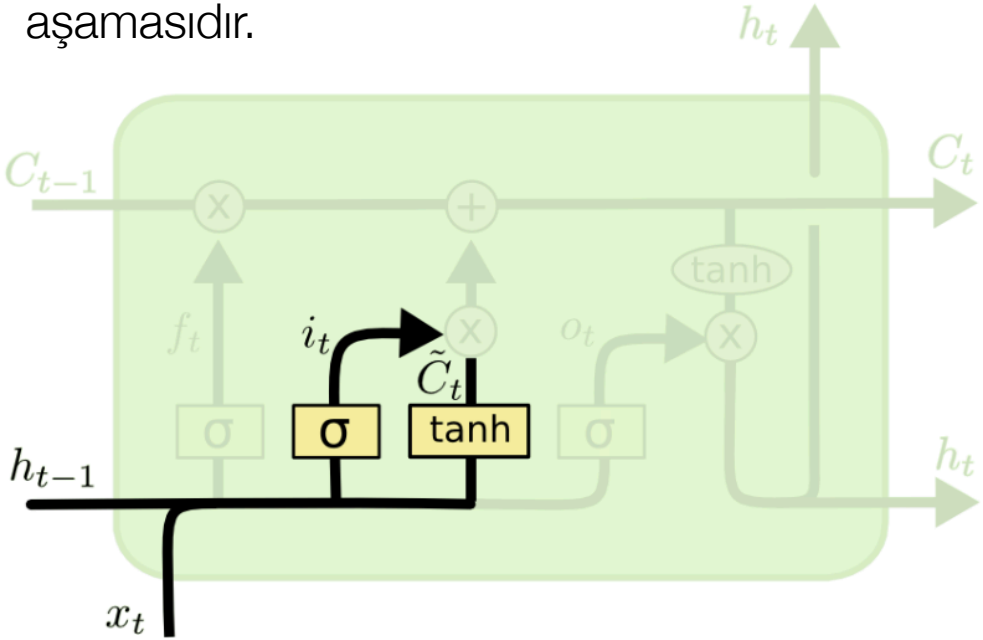
h_{t-1} ve x_t vektörlerine bakıp C_{t-1} hücre durumundaki her bir sayı için 0-1 arasında bir sayı üretir.

1 --> “bu bilgiyi tamamen sakla”
0 --> “bu bilgiyi unut”

Resim: C. Olah

LSTM : Girdi Kapısı ve hücre durumu

Bundan sonraki adım, hangi yeni bilginin hücre durumunda saklanacağına karar verme aşamasıdır.



Resim: C. Olah

Girdi kapısı bir sigmoid katmanıdır ve hangi değerleri güncelleyeceğimize karar verir.

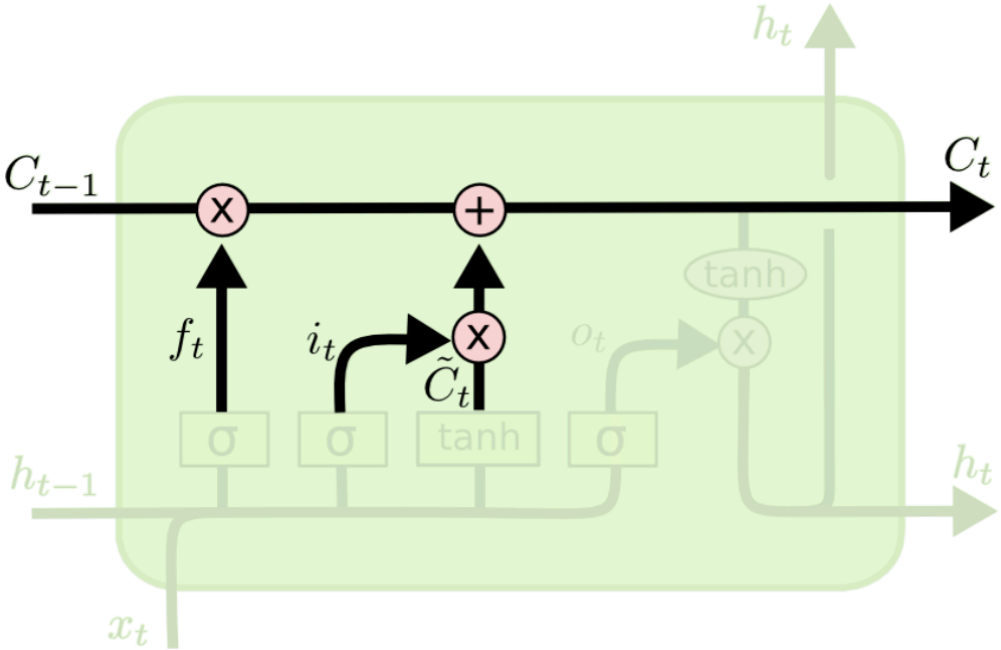
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

Bir tanh katmanı bir önceki zamandan gelen bilgileri ve input vektör bilgilerini birleştirerek, şu anki hücre durumuna eklenebilecek aday bilgi vektörünü oluşturur.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

LSTM : Girdi Kapısı ve hücre durumu

Yeni hücre durumu şu şekilde güncellenir:



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

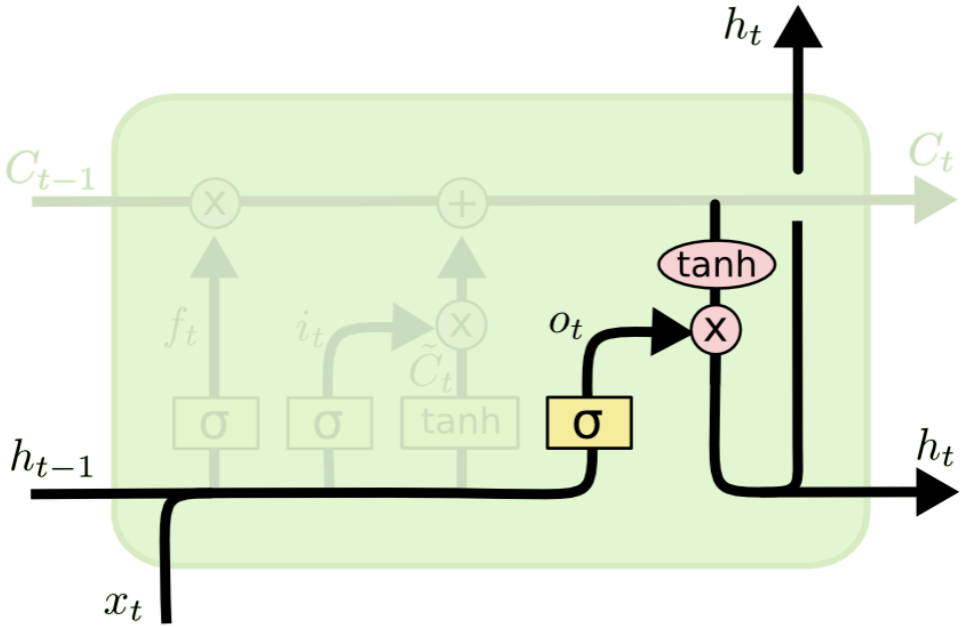
Eski durumu ft ile çarparak unutmaya karar verilen bölümleri unuturuz.

Daha sonra yeni aday değerleri, her birini ne kadar güncellemek istediğimize bağlı olarak ekleriz.

Resim: C. Olah

LSTM : Çıktı

Son olarak, üretilecek çıktıya karar vermemiz gerekiyor:



Resim: C. Olah

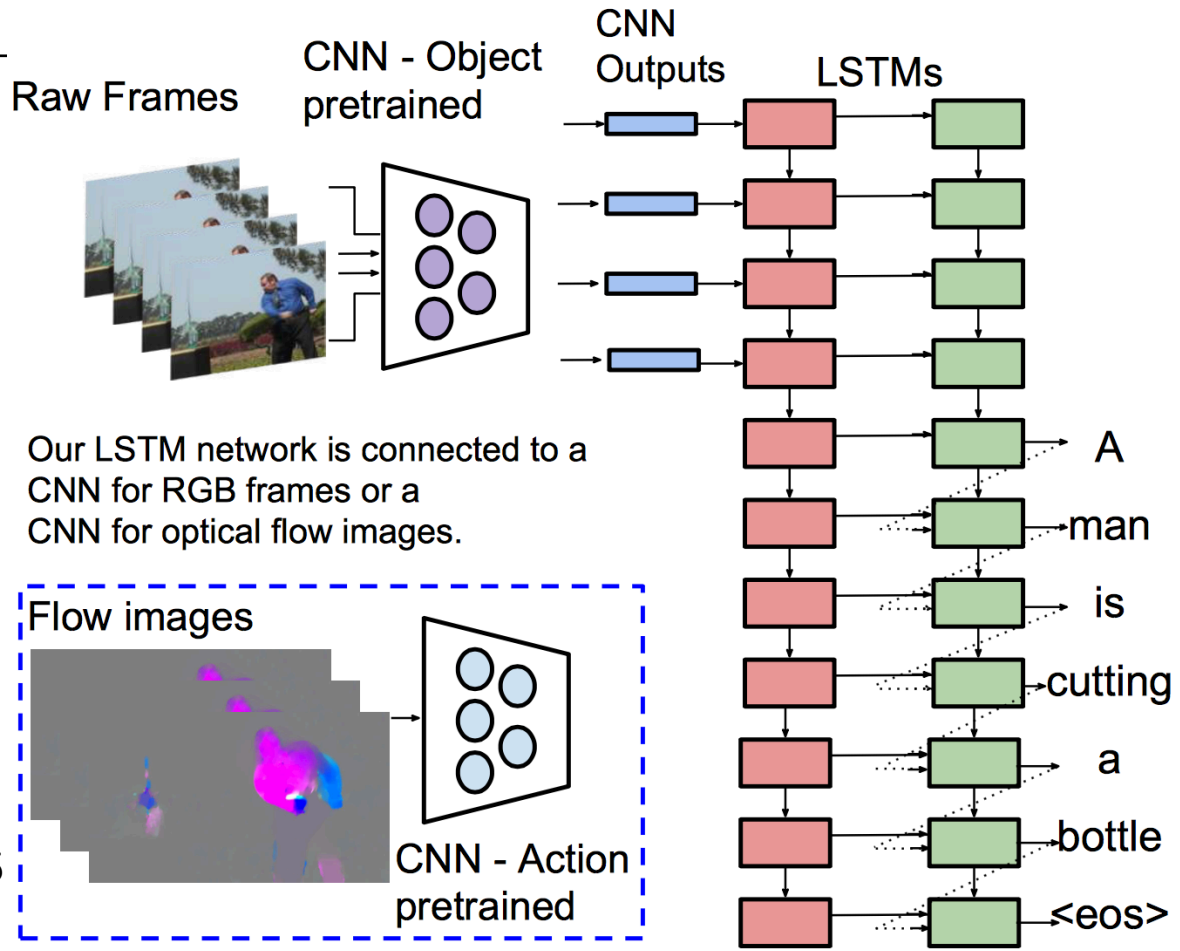
Önce, hücre durumunun hangi bölümünü güncelleyeceğimize karar vermek için bir sigmoid katmanı kullanılır:

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

Hücre durum vektörünü tanh'ten geçiririz (değerleri [-1,1] arasına çekme için) ve sadece karar verdiğimiz bölümleri çıktı vermek için üsteki sigmoid kapısı ile çarparız.

$$h_t = o_t * \tanh\left(C_t\right)$$

Uygulamalar: Videoları doğal dille anlatma

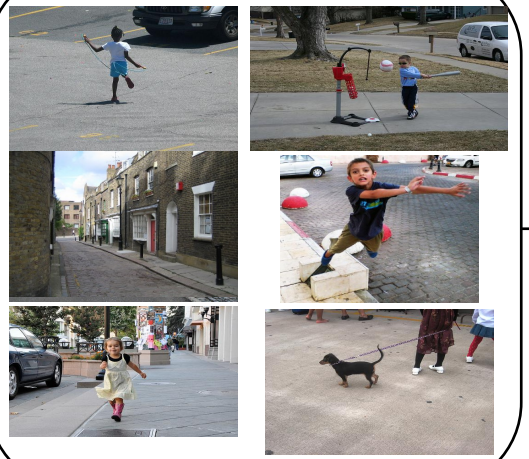


Venugopalan et al. ICCV 2015

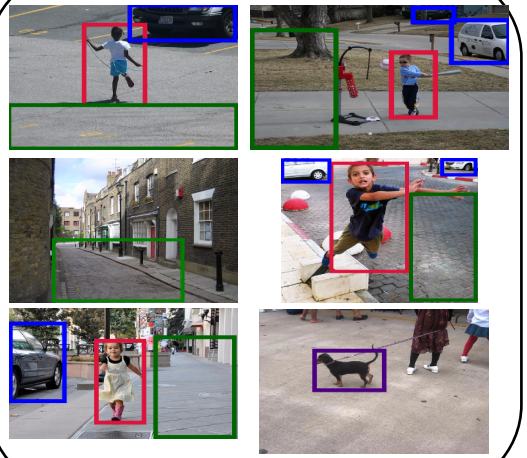
Belirgin alanlar ile İmge Altyazılama

Sorgu imgesi

Global analiz



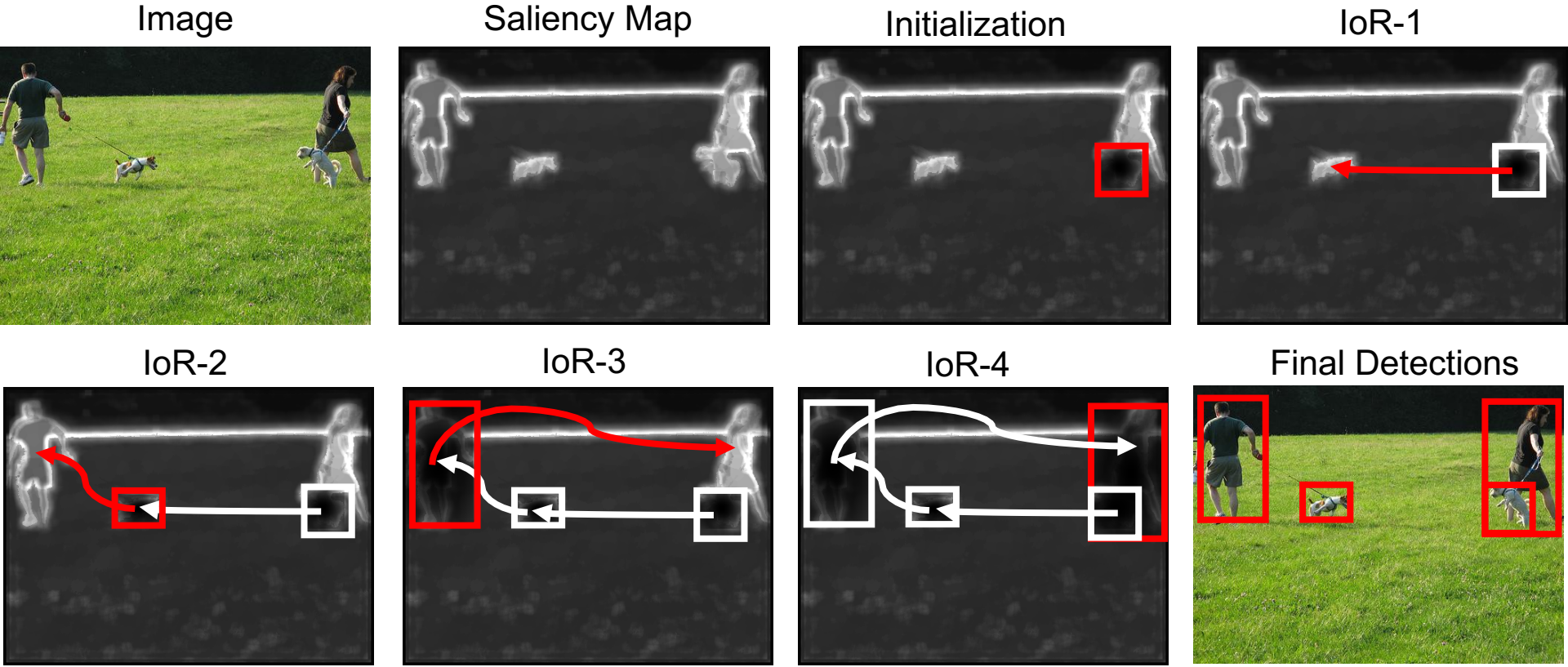
Lokal analiz



A small child besides car is playing on the street

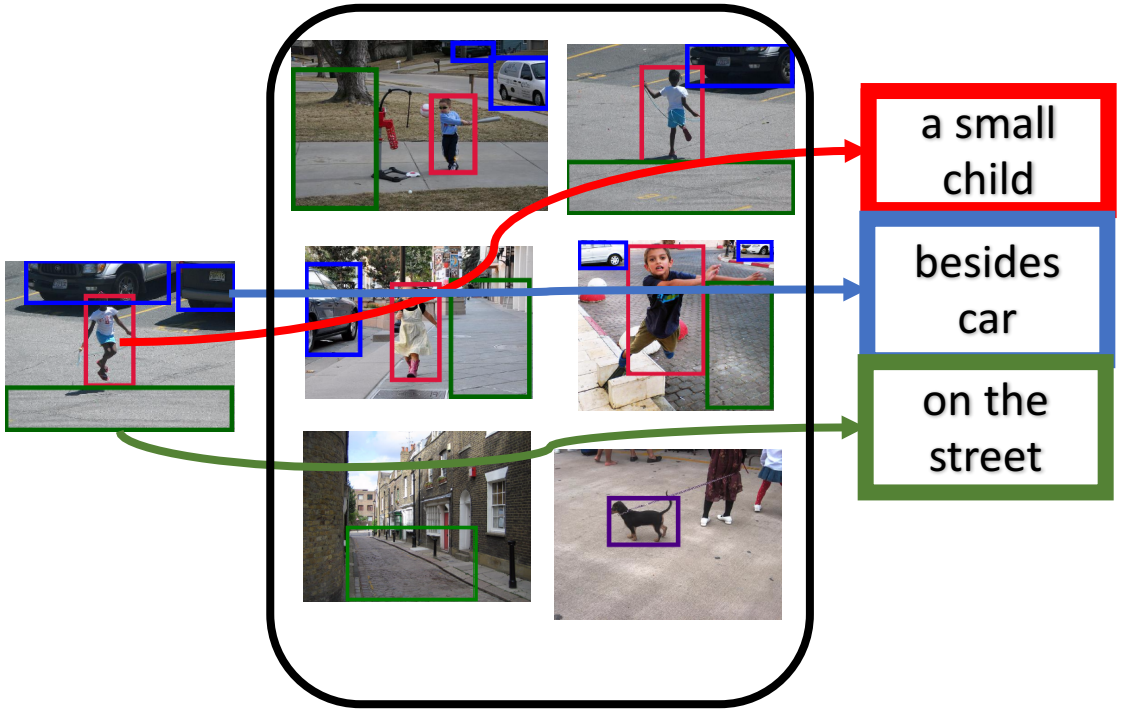
A boy is hitting a ball outside

Mert Kılıçkaya, Burak Kerim Akkuş, Ruket Çakıcı, Aykut Erdem, Erkut Erdem, Nazlı İkizler Cinbiş, Ruken Çakıcı, Data-Driven Image Captioning via Salient Region Discovery, in submission.



Mert Kılıçkaya, Burak Kerim Akkuş, Ruket Çakıcı, Aykut Erdem, Erkut Erdem, Nazlı İkizler Cinbiş, Ruken Çakıcı, Data-Driven Image Captioning via Salient Region Discovery, in submission.

Önemli alt alanların bulunması



Object NPs

Actions VPs

Stuff PPs

Scene PPs

Probabilistic Language Models

Description

Mert Kılıçkaya, Burak Kerim Akkuş, Ruket Çakıcı, Aykut Erdem, Erkut Erdem, Nazlı İkizler Cinbiş, Ruken Çakıcı, Data-Driven Image Captioning via Salient Region Discovery, in submission.



Reranking w/ BoOP	A group of people hold hands on the beach	Dog catching bubbles on the grass	A bicyclist jumping with his bike in midair
HYBRID	A man wearing swimming shorts on a beach goes for a header with a soccer ball	A white dog is trying to play with a brown dog	A man is jumping in the air on his skateboard

Mert Kılıçkaya, Burak Kerim Akkuş, Ruket Çakıcı, Aykut Erdem, Erkut Erdem, Nazlı İkizler Cinbiş, Ruken Çakıcı, Data-Driven Image Captioning via Salient Region Discovery, in submission.

Nesne tanımayı iyileştirmek için doğal altyazıların kullanılması

A plane here flies!




Mert Kılıçkaya, Nazlı İkizler Cinbiş, Aykut Erdem, Erkut Erdem, Leveraging captions in the wild to improve object detection, THE 5TH WORKSHOP ON VISION AND LANGUAGE, in conjunction with ACL 2016.

Açıklamalar ve Doğal Altyazılar

Açıklama

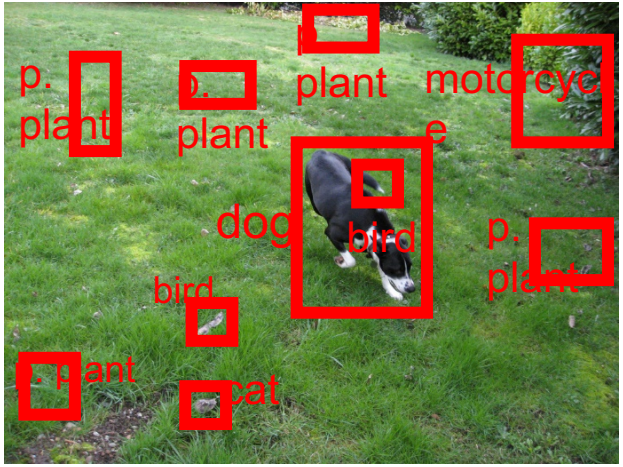
-There is a **bottle** on the leaves
- Someone placed a **plane** in a **bottle** standing on the ground
- This is an image of a green **bottle** surrounded by the grass
- A poetic scene of a **plane** flying in a huge **bottle**

Doğal Altyazı

- A **plane** here flies!

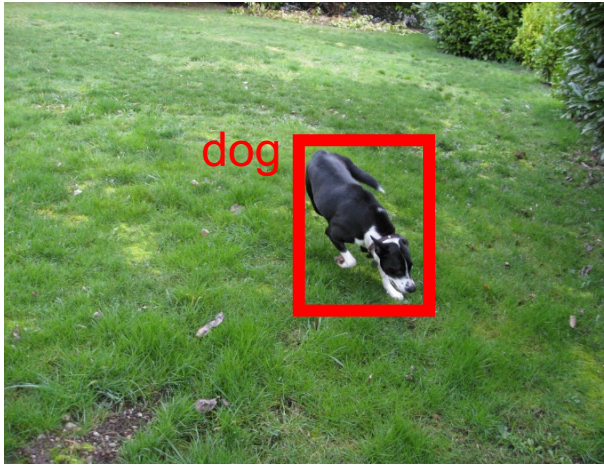
Nesne tanımayı iyileştirmek için doğal altyazıların kullanılması

Detections



Ordonez et al.*: *Motorcycle*

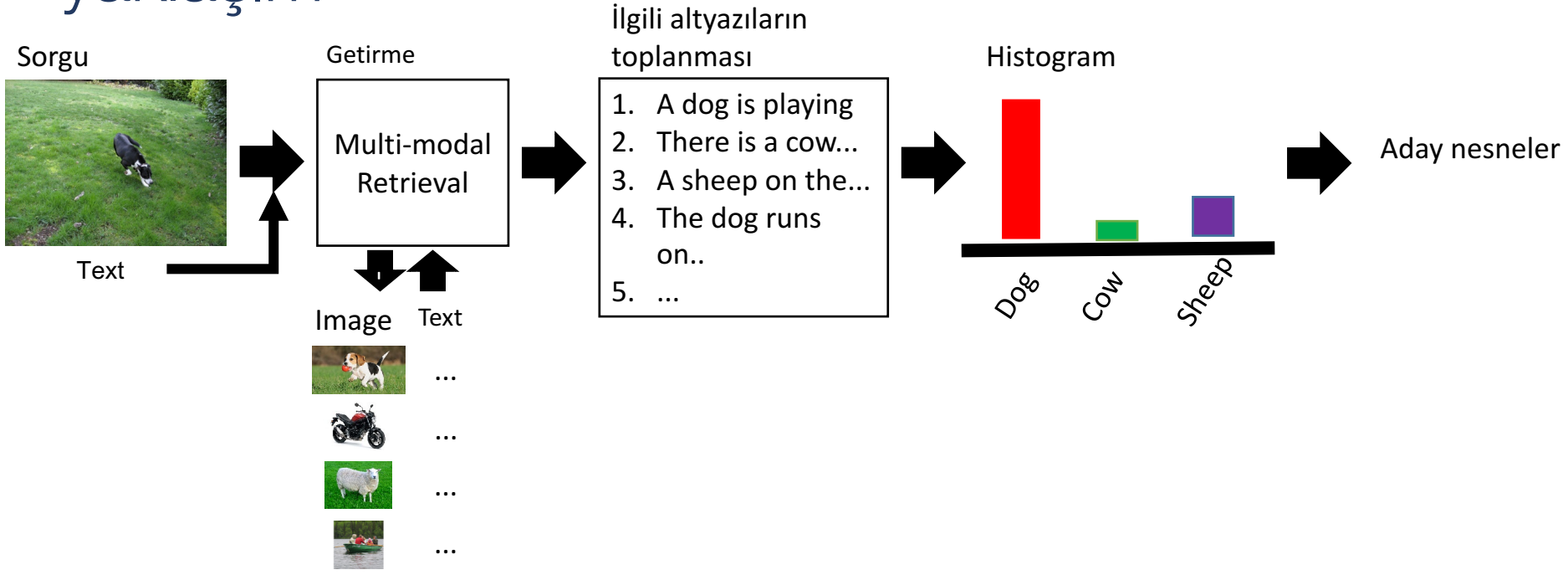
Our approach: *Dog*



Caption: Probably in pursuit of a **motorcycle** going up on the road past our house, or similar

* Ordonez, Vicente, et al. "Large scale retrieval and generation of image descriptions." International Journal of Computer Vision (2015): 1-14.

Aday nesnelerin bulunması için veri odaklı bir yaklaşım



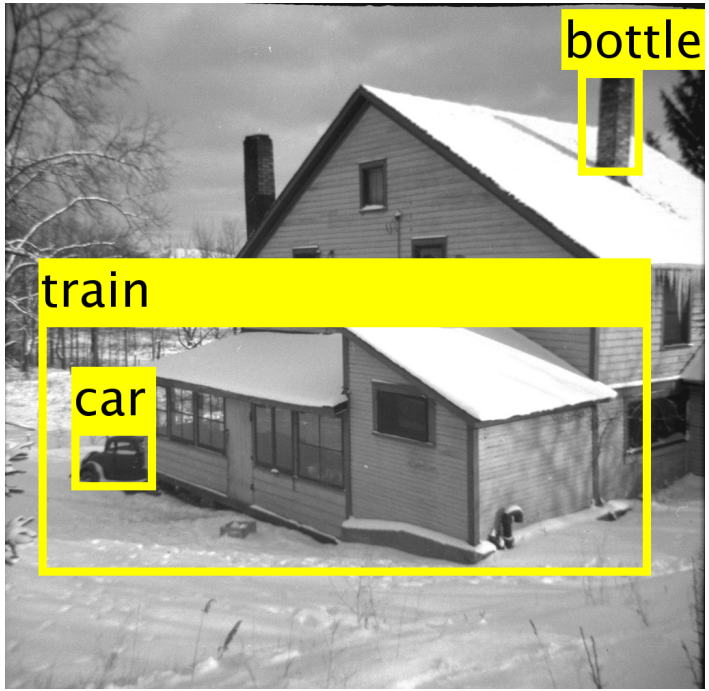
[1]

Önerilen Yöntem

[1] Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

55

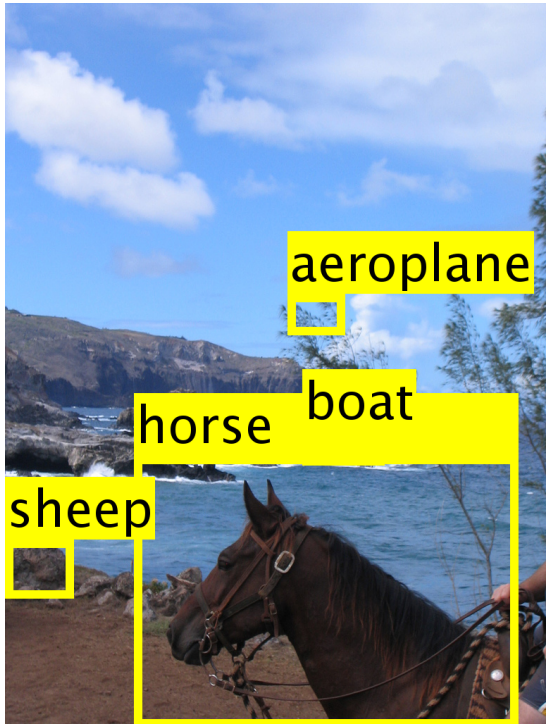
[1]



Önerilen Yöntem

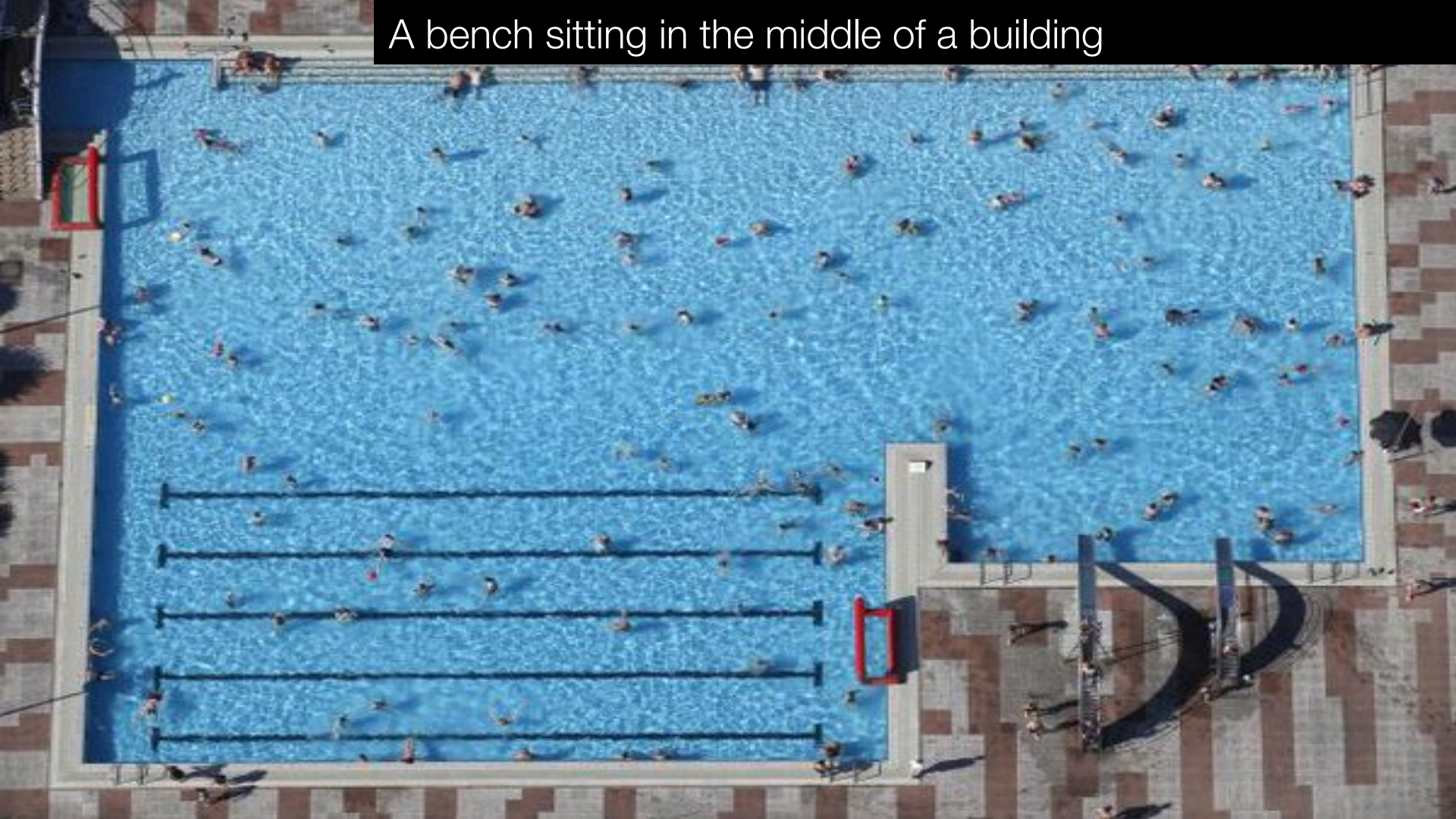
[1] Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[1]

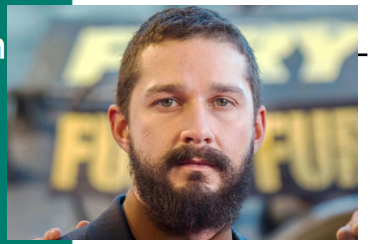


Önerilen Yöntem

[1] Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.


A bench sitting in the middle of a building

I think it's a group of people posing for a picture and they seem 😁😁😐😁😁😁😁😁😁😁😁😁😁. I am 96% sure that's Shia LaBeouf



Bir kadın ve bir adam sarılmışlar.

https://www.captionbot.ai

Görü ve Dil için Derin Öğrenme

Aykut Erdem, Erkut Erdem, Nazlı İkizler Cinbiş



HACETTEPE UNIVERSITY
COMPUTER VISION LAB

@HUCVL

TÜBİTAK-COST 113E116