



Görseller için Kısa İçerik  
Üretmek için Yapay Zeka

- ❖ Web içerik yayınlıyorsak
- ❖ Görsellere açıklayıcı metin üretmek
  - sıkıcı bir iş,
  - yeterli kaynak,
  - zaman

- Web sitelerine erişilebilirliği artırmak önemlidir
- SEO (Search Engine Optimization) önemi
- Web developer ve içerik editörleri bazen SEO performansını unuttur veya görmezden gelir

## Image Alt Text

Doritos Tortilla Chips, Nacho Cheese, 1.75-Ounce Large Single Serve Bags (Pack of 64) Doritos

★★★★☆ 175 customer reviews | 9 answered questions



### About the Product

- Pack of 64, 1.75-ounce large single-serving bags (total of 112 ounces)
- Doritos tortilla chips with classic nacho-cheese flavor
- Made of whole corn; flavorings feature real cheddar cheese

```
,{"id":"defect-image-offensive","value":"Offensive or adult content"},  
{ "id":"defect-image-extra-items","value":"Shows additional items"},  
{ "id":"defect-image-not-clear","value":"Is not clear"}, {"id":"defect-  
other-image-issue","value":"Other"}]] data-metadata="IMAGE" data-  
feature-container-id="imageBlockEDPOverlay" data-custom-event-handler=  
"imageBlockEDPCustomEventHandler" data-display-name="Images" data-edit-  
data-state="imageBlockEDPEditData" data-position="0" data-resolver=  
"CQResolver"></span>  
<!-- Creating a custom overlay for image not available experience -  
-->  
▶<div class="variationUnavailable unavailableExp" style="display:  
none;">...</div>  
<div class="a-hidden" id="auiImmersiveViewDiv"></div>  
▼<ul class="a-unordered-list a-nostyle a-horizontal list maintain-  
height">  
::before  
<span id="imageBlockEDPOverlay"></span>  
▼<li class="image item itemNo0 maintain-height  
selected" style="cursor: pointer;">  
▼<span class="a-list-item">  
▼<span class="a-declarative" data-action="main-image-click" data-  
main-image-click="{ }">  
▼<div id="imgTagWrapperId" class="imgTagWrapper" style="height:  
500px;">  


https://alttext.ai/



Products ▾

Pricing

Support

Sign In

Start Free Trial

# AI Alt Text Generator

Automatically generate alt text for images in over 130 languages. Improve SEO and site accessibility. Integrations for WordPress, Shopify, web, CMS, and more.

Free Trial

How It Works →

Try it with  
your image



Click to select or drag and drop

JPG, WebP, GIF, PNG (max 10MB)

🔍 Choose File



Add custom SEO keywords

Optional

Free Trial

How It Works →



Try it with  
your image



Add custom SEO keywords

Optional

Generate Alt Text ↗

**A golden retriever puppy sitting on a white background.**

**A friendly golden retriever puppy sitting on a medium-size white background.**

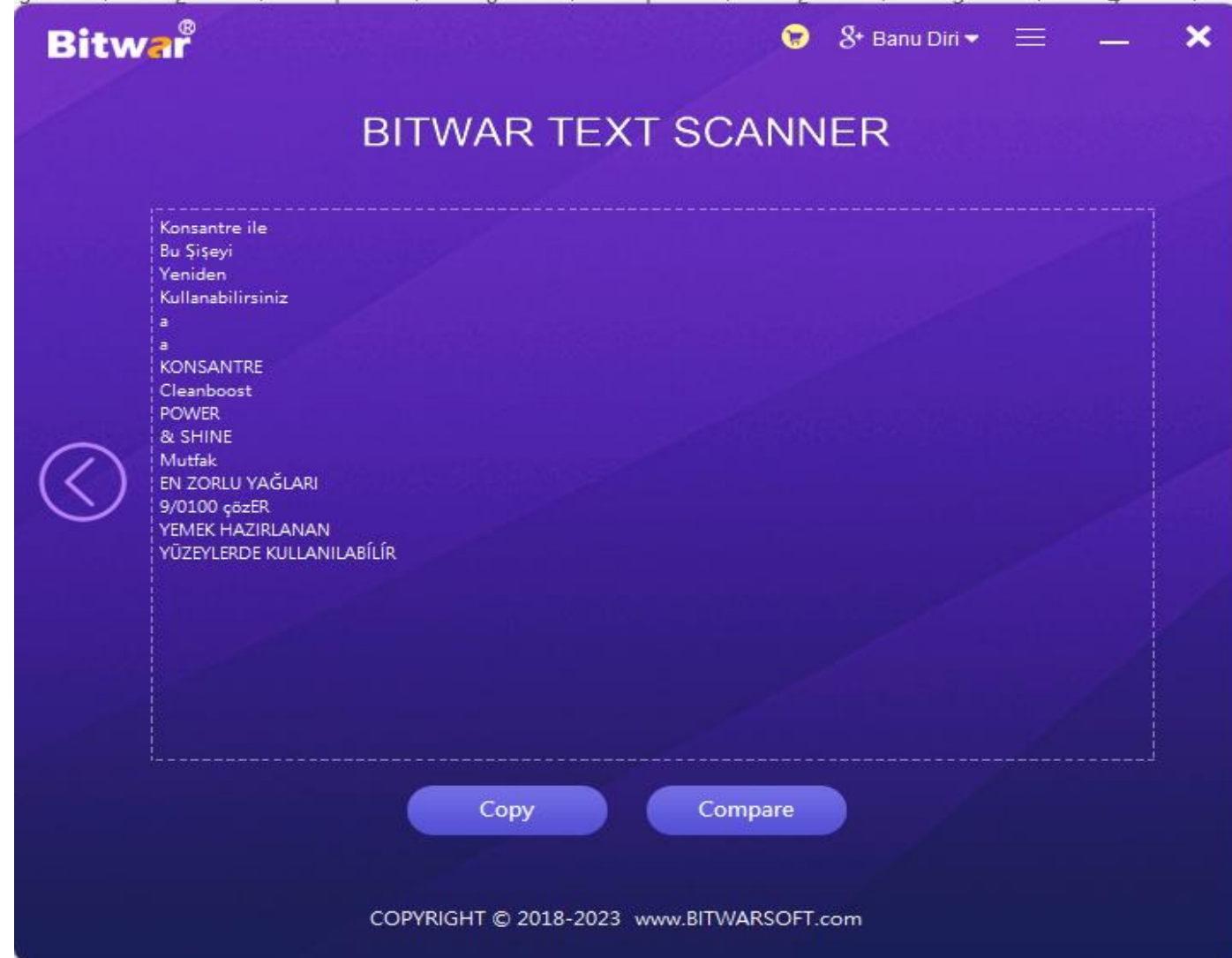
**Beyaz bir arka plan üzerinde oturan bir Golden Retriever köpek yavrusu**

**CEO anahtar  
kelimeler  
Scottish, friendly,  
medium size**



Görüntü üzerinden yazılı bir metni çıkarmak için OCR kullanılır

- ❖ Bitwar Text Scanner
- ❖ Tesseract
- ❖ .....



## Görüntüden Metin Üretmek – Image Captioning

Görüntünün içeriğinin otomatik olarak belirlenmesi ve doğal dil yardımıyla insan benzeri bir açıklamanın otomatik olarak çıkarılması, görüntüden metin üretilmesi olarak adlandırılır.

Bir görüntünün metinsel tanımını oluşturma süreci :

- görüntüdeki nesneleri tanıma
- anlamsal ilişkileri çıkarma
- arka plan sahnesini anlama
- bilgiyi sözdizimsel olarak doğru cümlelere dönüştürme

**AMAÇ:** Verilen görüntüyü anlamsal olarak en iyi tanımlayan cümleyi üretmek



## Veri seti : Microsoft Common Objects in Context (MS COCO)

MS COCO, nesne segmentasyonu, bağlam tanınması ve görüntü altyazısı oluşturma çalışmalarında kullanılır.

- 80 adet Nesne kategorisi (kişi, araba, köpek vb.)
- 91 adet Alan kategorisi (gökyüzü, sokak, deniz vb.)
- 160.000 (160K)'dan fazla etiketli resim olup, toplam 330K görüntüye sahiptir
- Her görüntü İngilizce dilinde yaklaşık 5 altyazı cümle ile etiketlenmiştir (Amazon Mechanical Turk çalışanları tarafından etiketlenmiştir).

**IAPR TC-12, Flickr8K, Flickr30K**



a horse with a half white and half black face.  
a horse in a pen stands on the side in partial shade.  
a horse standing on a dirty floor next to a wooden fence.  
an image of a horse that is inside of a coral  
a gray horse wit a white muzzle is standing in a fenced enclosure.

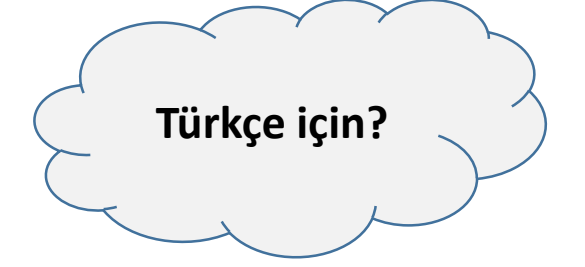
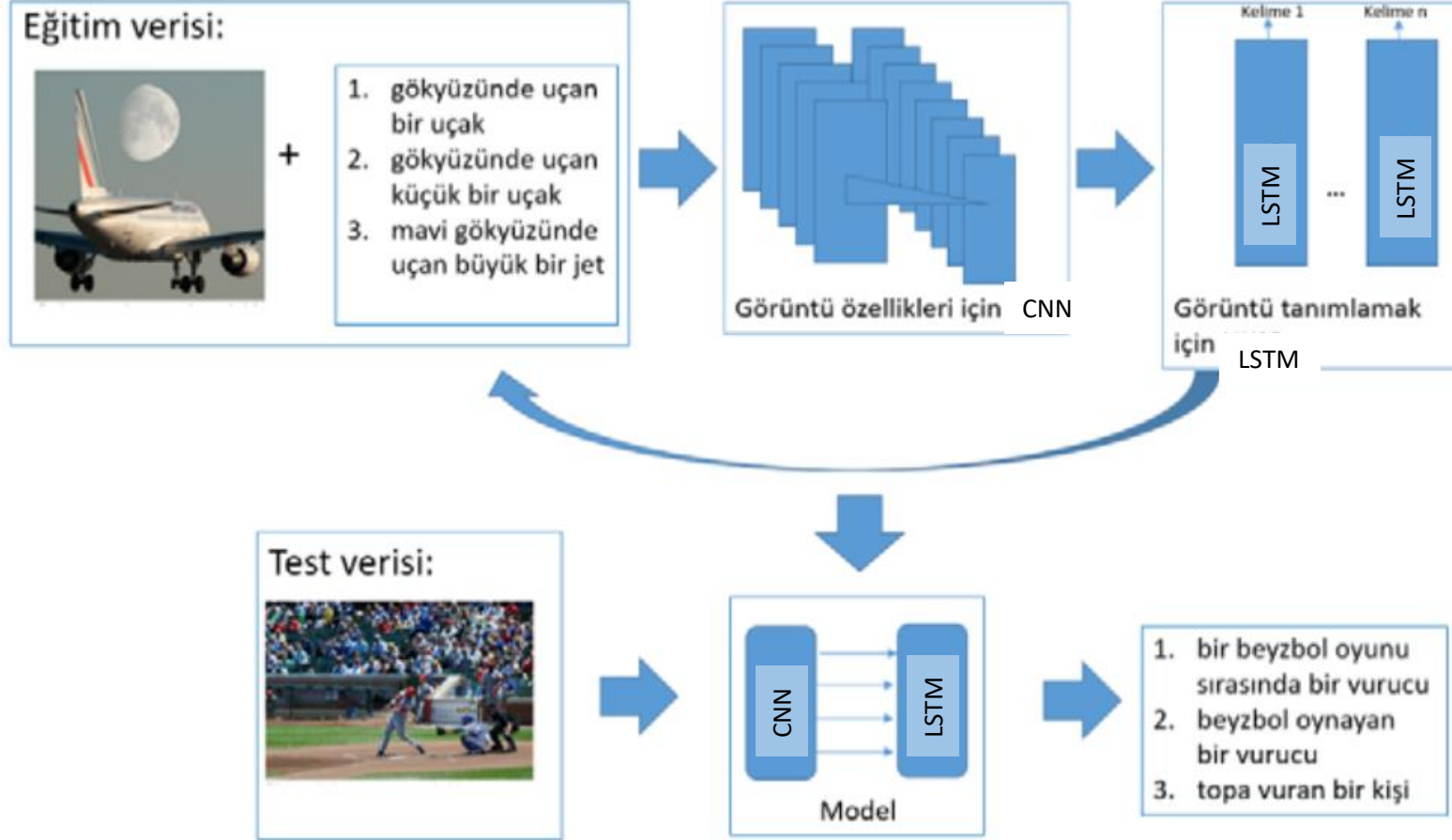
## Makine öğrenmesi ve Derin öğrenme yöntemleri kullanılmıştır

| Yazar                  | Veriseti         | Yöntem       | Sonuçlar                  |
|------------------------|------------------|--------------|---------------------------|
| Bai ve ark. [15]       | MC-COCO          | MGCN         | BLUE2=0.567 METEOR=0.281  |
| Kilickaya ve ark. [16] | Pascal cümleleri | Im2Text      | BLUE1=0.11                |
| Lu ve ark. [17]        | MSCOCO+ArtCap    | Faster R-CNN | BLUE1=0.508 METEOR=0.1317 |
| Yang ve ark. [18]      | HS-COCO          | HCCM         | BLUE1=0.839 METEOR=0.304  |
| Agrawal ve ark. [19]   | MS-COCO          | Inception V3 | -----                     |
| Li ve ark. [20]        | MS-COCO          | CNN+RNN      | -----                     |

## Literatür :

1. C. Bai, A. Zheng, Y. Huang, X. Pan, and N. Chen, "Boosting convolutional image captioning with semantic content and visual relationship," *Displays*, vol. 70, p. 102069, Dec. 2021 doi: 10.1016/j.displa.2021.102069.
2. M. Kılıçkaya, E. Erdem, A. Erdem, N. İ. Cinbiş, and R. Çakıcı, "Data-driven image captioning with meta-class based retrieval," in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, Apr. 2014, pp. 1922–1925. doi:10.1109/SIU.2014.6830631.
3. Y. Lu, C. Guo, X. Dai, and F.-Y. Wang, "Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training," *Neurocomputing*, vol. 490, pp. 163–180, Jun. 2022 doi: [10.1016/j.neucom.2022.01.068](https://doi.org/10.1016/j.neucom.2022.01.068).
4. Z. Yang, P. Wang, T. Chu, and J. Yang, "Human-Centric Image Captioning," *Pattern Recognition*, vol. 126, p.108545, Jun. 2022, doi: [10.1016/j.patcog.2022.108545](https://doi.org/10.1016/j.patcog.2022.108545).
5. V. Agrawal, S. Dhekane, N. Tuniya, and V. Vyas, "Image Caption Generator Using Attention Mechanism," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2021, pp. 1–6. doi:10.1109/ICCCNT51525.2021.9579967.
6. J. Li, N. Xu, W. Nie, ve S. Zhang, "Image Captioning with multi-level similarity-guided semantic matching", *Visual Informatics*, c. 5, sy 4, ss. 41-48, Ara. 2021 doi: [10.1016/j.visinf.2021.11.003](https://doi.org/10.1016/j.visinf.2021.11.003).

\* Geliştirilecek sistem, görüntülerden özellikleri çıkarmak için CNN mimarisine sahip kodlayıcı ve ilgili alt metni oluşturmak için RNN mimarisine sahip kod çözücünden oluşmaktadır. (Türkçe için yapılmıştır)



### Çalışmada

- yığın boyutu/batch size 128
- gömme boyutu/embedding size 512
- LSTM gizli katman boyutu 512
- öğrenme oranı/learning rate 1e-3

|                | Train | Validation | Test | Total |
|----------------|-------|------------|------|-------|
| Görüntü sayısı | 83K   | 41K        | 41K  | 165K  |



İYİ



Gökyüzünde uçan büyük bir ticari uçak.



Bir beyzbol oyunu sırasında bir vurucu, kaleci ve hakem.



Bir uçak penceresinden bir dağ manzarası.

ORTA



Bir oyuncak ayı tutarak yatakta yatan bir adam.



Küçük bir çocuk bir diş fırçası ile dişlerini fırçalıyor.



Bir zürafa bir ağacın yapraklarını yiyor.

KÖTÜ



Turuncu ve beyaz bir kedi bir battaniye üzerinde yatıyor.



Siyah bir ayı bir ağaç dalı üzerinde oturuyor.

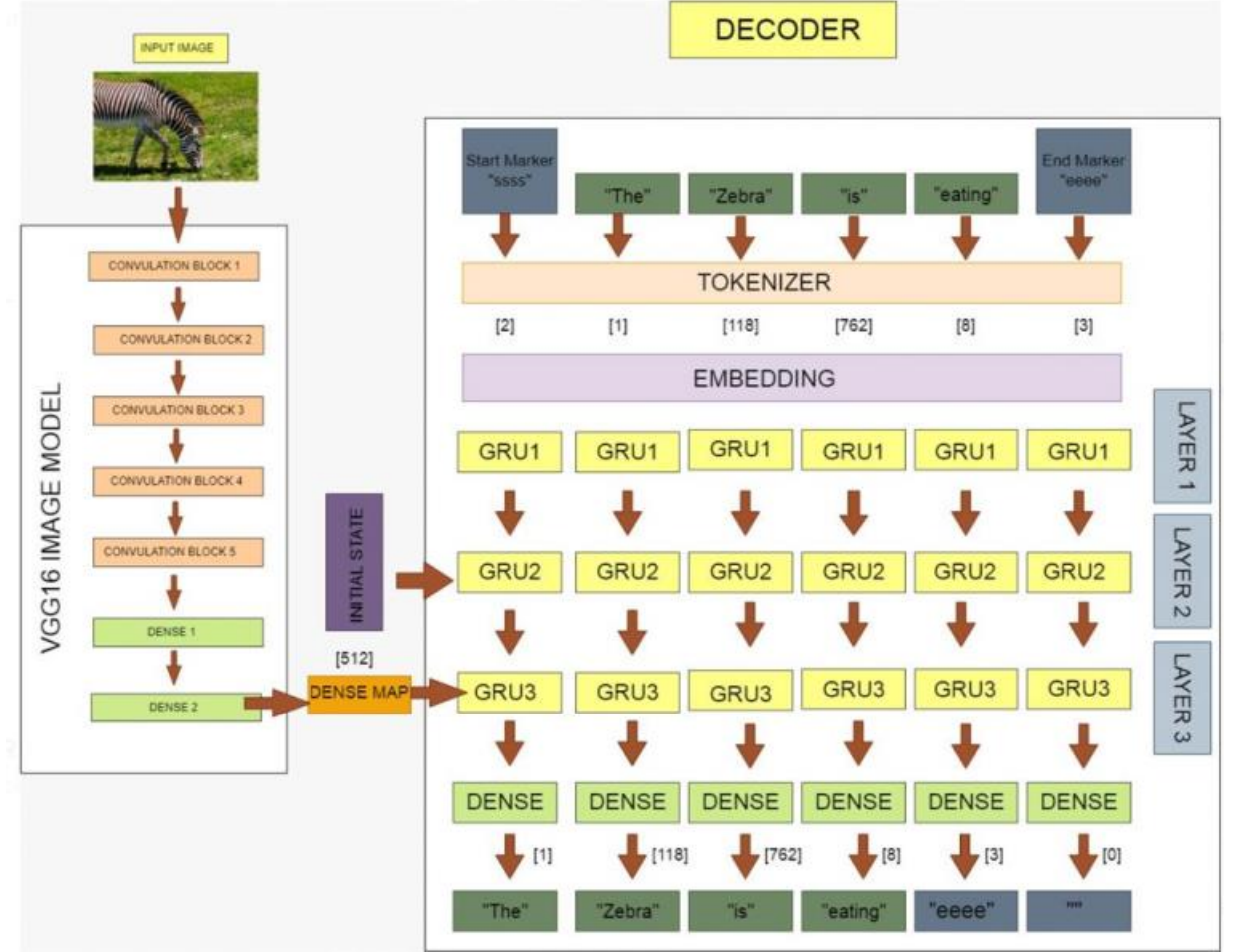


Ahşap bir masa üzerinde bir elma ve bir portakal.

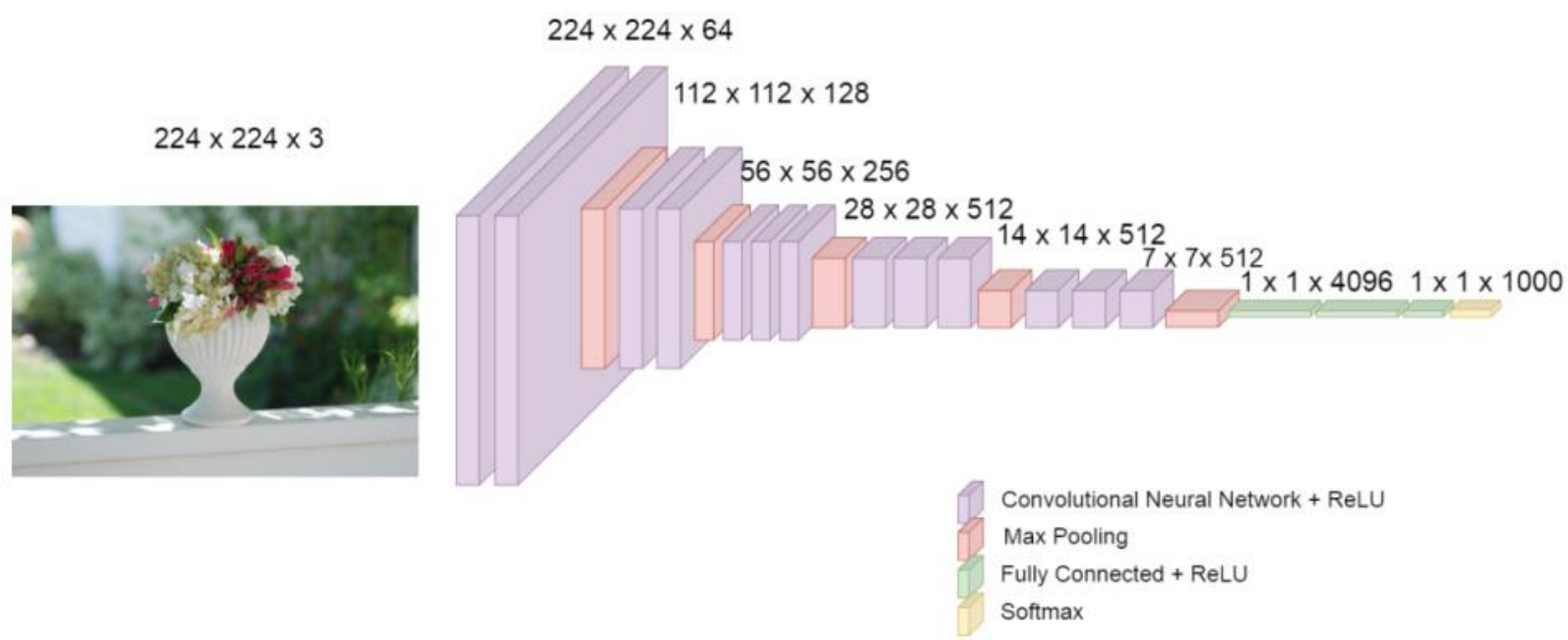
| * Algorithm | Model-1 |
|-------------|---------|
| BLEU-1      | 0,297   |
| BLEU-2      | 0,164   |
| BLEU-3      | 0,076   |
| BLEU-4      | 0,035   |
| ROUGE_L     | 0,272   |
| METEOR      | 0,129   |
| CIDEr       | 0,528   |

\*\* Bu çalışmada, görüntülerin özniteliklerinin çıkarılmasında CNN mimarisine sahip bir kodlayıcı, altyazıların oluşturulmasında da RNN mimarisine sahip bir kod çözücü kullanılmıştır.

- MS-COCO veri kümesi (İngilizce için hazır) Eğitim için 118.287, test için 5000 olmak üzere toplamda 123.287 görüntü
- Görüntüleri tanımda VGG16 veri modeli kullanılmıştır.
- VGG16 modeline görüntü verildikten sonra 5 convolution bloğu ve 2 dense katmanından geçirilerek düşünce vektörü (thought vector) üretilmiştir.
- Görüntünün içeriğini açıklayan thought vektörü ve doğru cümle, kod çözücüye girdi olarak verilmiştir.
- Kod çözücüye verilen cümlelerin başına başlangıç belirteci, sonuna ise bitiş belirteci eklenmiştir.
- Kod çözücü, çıktı olarak kelimeleri üretmektedir.



GRU (Gated Recurrent Units) → Kapı Özyinelemeli Geçitler  
LSTM'ler ve GRU'lar vanishing gradient problemi çözümü için oluşturulmuştur



- VGG16, görsel nesne tanımada kullanılan bir CNN mimarisidir.
- 16 sayısı, ağırlıklarla birlikte 16 katmanı olduğunu söyler.
- 138 milyon parametreye sahiptir.
- Eğitim sırasında, ImageNet veri kümesi  $224 \times 224$  RGB görüntü içermektedir.  $(224,224,3)$  tensor bulunur.
- 13 evrişim katmanına, 5 max-pooling katmanına, üç tam bağlantılı katmana ve softmax katmana sahiptir.
- Aktivasyon fonksiyonu olarak ReLU kullanmaktadır.





A giraffe eating food from the top of the tree.  
 A giraffe standing up nearby a tree.  
 A giraffe mother with its baby in the forest.  
 Two giraffes standing in a tree filled area.  
 A giraffe standing next to a forest filled with trees.



A zebra grazing on lush green grass in a field.  
 Zebra reaching its head down to ground where grass is.  
 The zebra is eating grass in the sun.  
 A lone zebra grazing in some green grass.  
 A zebra grazing on grass in a green open field.

Az nesne içeren görüntüler için metin üretme sonuçları



An unusual looking red bus going down a road.  
 A bus driving down the street next to some buildings and trees.  
 A red tram carrying a group of people on a tour.  
 A red double decker vehicle driving down a street.  
 A big red bus that is driving down the road.



A living room filled with furniture on top of a hard wood floor.  
 A living room with big couches and a ceiling fan.  
 A living room filled with couches chairs and a water cooler sitting on a hard wood floor.  
 Living room with wood floor TV couches and chair.  
 A very big nice-looking room with a bright window.



A watery glass jar full of blooming flowers.  
 A vase filled with yellow and white flowers.  
 A glass jar is filling with white and yellow flowers.  
 A glass vase is holding a bunch of flowers.  
 Tall glass mug shaped vase with yellow flowers.

Çok nesneli görüntüler için metin üretme sonuçları

| Algoritma | Model Performansı |
|-----------|-------------------|
| BLUE-1    | 0,641             |
| BLUE-2    | 0,420             |
| BLUE-3    | 0,334             |
| BLUE-4    | 0,185             |
| METEOR    | 0,056             |
| ROUGE_L   | 0,117             |