



# Derin Öğrenme Yöntemleri Yardımıyla Görüntülerin Otomatik Konumlu Betimlenmesi ve Türkçe Alt Yazı Oluşturulması için Yeni Veri Kümesi ile bir Yöntem Önerisi

## A Method Proposal for Creation Automatic Turkish Directional Depicted and Captioning of Images with Deep Learning Methods and the New Turkish Dataset

Esin ERGUVAN ETGİN  
Maltepe Üniversitesi  
Bilgisayar Mühendisliği Bölümü  
İstanbul, Türkiye  
eerguvans@gmail.com  
ORCID: 0000-0002-2607-6076

Erdal GÜVENOĞLU  
Maltepe Üniversitesi  
Bilgisayar Mühendisliği Bölümü  
İstanbul, Türkiye  
erdalguvenoglu@maltepe.edu.tr  
ORCID: 0000-0003-1333-5953

### Öz

Günümüzde görüntü işlemenin yaygın olarak kullanılması ile birlikte görüntülerin otomatik olarak betimlenmesi ve alt yazı oluşturulması önemli görülmektedir. Bu çalışma ile Türkçe otomatik konumlu betimleme ve alt yazı oluşturulması amaçlanmıştır. Bu çalışmada, MS-COCO veri kümesinden 1500 görüntü seçilmiştir. Seçilen her bir görüntü için görüntüde saptanan bazı nesnelerin birbirlerine göre konumlarını içerecek şekilde alt yazılar oluşturulmuş ve yeni bir veri kümesi elde edilmiştir. Bu veri kümesi kullanılarak otomatik konumlu betimleme ve alt yazılama için bir yöntem önerilmiştir. Bu yöntem için elde edilen en iyi altı model otomatik alt yazı oluşturma için seçilmiştir. Deneysel sonuçlar önerilen yöntemin Türkçe otomatik konumlu betimleme ve alt yazı oluşturma için başarılı sonuçlar ortaya koyduğunu göstermektedir.

**Anahtar sözcükler:** Görüntü alt yazılama, Bilgisayarlı görü, Doğal dil işleme, FasterRCNN, ResNet, GRU.

### Abstract

With image processing is widely used today, to be depicted and automatic captioning of images is of great importance. With this study, it is aimed to make an automatic directional depicted and captioning of the objects in the images in Turkish language. In this study, 1500 images were selected from the Microsoft Common Objects in Context (MS-COCO) dataset. For

each selected image, subtitles were created to include the relative positions of some objects detected in the image, and a new data set was obtained. Using this dataset, a method has been proposed for automatic to be directional depicted and captioning of images and the best six models obtained for this method were selected. Experimental results show that of the proposed method show successful results for automatic to be directional depicted and captioning of images in Turkish language.

**Keywords:** Image captioning, Computer vision, Natural language processing, FasterRCNN, ResNet, GRU.

### 1. Giriş

Bilgisayarlar ve bilgisayar sistemleri, artık dünyamızın vazgeçilmez bir parçasıdır. Günümüzde, bilgisayarlar hem olaylar hakkında karar verebilmekte hem de olaylar arasındaki ilişkileri öğrenebilmektedir. Bilgisayarları bu özellikler ile donatan ve bu yeteneklerin gelişmesini sağlayan çalışmalar "yapay zekâ" çalışmaları olarak bilinmektedir. İlk defa 1950'li yıllarda ortaya atılmıştır. Yapay zekâ, makine öğrenmesi ve derin öğrenme arasında iç içe bir ilişki söz konusudur. Derin öğrenme, çok katmanlı bir yapay sinir ağıdır. Veriden doğrudan özellik ve görevleri öğrenen bir makine öğrenmesi tekniğidir. İşlenecek veri görüntü, metin ya da ses dosyası olabilmektedir.

Günümüzde derin öğrenme pek çok alanda başarılı bir şekilde yaygın olarak kullanılmaktadır. Bu gelişme, internet sayesinde büyük boyutlarda verinin sayısal ortamda üretilir ve saklanır hale gelmesi, aynı zamanda hızlı GPU'lar ile işlem gücünün

artmasından dolayı gerçekleşmiştir. Çok katmanlı bir yapay sinir ağı olan derin öğrenme yöntemleri ile yüksek katman sayısı sayesinde çok daha karmaşık olan sorunlar günümüzde kolaylıkla çözülebilmektedir.

Görüntü işleme, bilgisayarlı görü ve derin öğrenme gibi bilim dallarındaki gelişmelerle birlikte, görüntü alt yazılama ve görüntüyü betimleme üzerine çalışmalar da hız kazanmıştır. Görüntülerin otomatik betimlenmesi ve alt yazı çıkarılması konusu, kaynaklarda daha yeni incelenmeye başlanmıştır. Görüntüyü betimleme, görüntü tanıma ve doğal dil işleme algoritmaları birlikte kullanılmaktadır. Bundan dolayı zorlu bir çalışma alanı olarak kabul edilmektedir [1-4].

Makalede görüntülerin otomatik alt yazılama için yeni bir veri kümesi oluşturulmuş ve bir yöntem önerilmiştir. İkinci bölümde ilgili çalışmalara yer verilmiştir. Üçüncü bölümde veri kümesinin hazırlanması ve özellikleri ele alınmıştır. Dördüncü bölümde görüntülerin otomatik alt yazılama için faydalanılan ve önerilen yöntemler anlatılmıştır. Beşinci bölümde önerilen yöntemin başarımlarını değerlendirmeleri ve görüntü çıktılarını yer verilmiştir. Altıncı ve son bölümde sonuçlar ele alınmıştır.

## 2. İlgili Çalışmalar

Bilgisayarlı görü ve doğal dil işleme modelleri görüntülerin tasvir edilmesinde birlikte kullanılmaktadır. Bu birliktelik derin öğrenme çalışmalarında da sıklıkla tercih edilmektedir. Literatürde konu ile ilgili çalışmalardan bazıları aşağıda ele alınmıştır.

Antol ve ark. [5] tarafından görsel soruların yanıtlanması üzerine bir çalışma yapılmıştır. Görsel sorular, arka plan ayrıntıları ve altta yatan bağlam dâhil olmak üzere, görüntünün farklı alanları seçilmiştir. Makalede, genel görüntü alt yazıları üreten bir sistemden farklı olarak, görüntünün ve karmaşık akıl yürütmenin anlaşılmasına odaklanılmıştır. Lin ve ark. [6] makalesinde günümüzde yapay ajanların olgusal soruları cevaplayabildiği, fakat sağlıklı akıl yürütmeyi gerektiren yetenekten yoksun oldukları tezi ele alınmıştır. Görüntülerden boşluğu doldurma ve görsel yorumlama soruları üzerinde çalışılmıştır. Bu sorular yanıtlanırken, metnin arkasındaki sahne "hayal edilmiş" ve metinsel ipuçlarına ek olarak "hayal edilen" sahnelerden de görsel ipuçları kullanılmıştır. Vinyals ve ark. [7] İngilizce otomatik alt yazı oluşturma için iyileştirilmiş NIC (Nöral Görüntü Alt yazısı, Neural Image Caption) modelini önermişlerdir. Görüntü tespiti için 22 katmanlı GoogleLeNet, İngilizce dil üretimi için LSTM (Uzun Kısa-Dönemli Bellek Modeli, Long Short-Term Memory) modeli ve veri kümesi olarak SBU, Flickr, Pascal ve MS-COCO kullanılmıştır. Çalışmada daha ayrıntılı ve isteğe yönelik tanımlamalar amaçlanmıştır. Agrawal ve ark. [8] tarafından İngilizce görüntü alt yazısı oluşturma için NOCAPS (Ölçekte Yeni Nesne Alt Yazılama, Novel Object Captioning at Scale) modeli önerilmiştir. NOCAPS içindeki görüntüler Open Images V4 doğrulama ve sına kümelerinden alınmıştır. Veri kümesi 15.100 görüntüyü açıklayan 166.100 insan tarafından oluşturulan alt yazıdan oluşmaktadır. Open Images, COCO veri kümesinden çok daha fazla sınıf içermektedir, sına k görüntülerinde görülen yaklaşık 400 nesne sınıfı, hiç veya çok

az ilişkili eğitim başlığına sahiptir. Wu ve ark. [9] tarafından, dil dizisi modelini nesne açıklamalarından tamamen ayırabilen, bir ayrılmış yeni nesne alt yazısı (DNOC- Decoupled Novel Object Captioner) modeli önerilmiştir. Yeni nesne sınıfları üzerinde yapılan çalışmalarda, mevcut önceden eğitilmiş alt yazı modelleri yetersiz kalmaktadır. DNOC modeli bu soruna çözüm olarak tasarlanmıştır. Önerilen modelde kullanılan iki bileşen ile, dil dizisi modeli yeni nesne tanımlarından ayrılmıştır. Çalışmada nesne saptama modeli oluştururken Faster R-CNN (Daha Hızlı Bölge-Tabanlı Evrişimli Sinir Ağları Modeli, Faster Region-based Convolutional Neural Network) ve Inception-ResNet-v2 (Artık Ağlar, Residual Network) yöntemleri kullanılmıştır. Venugopalan ve ark. [10], çalışmada mevcut görüntü alt yazısı veri kümelerinde bulunmayan çok sayıda nesne sınıflarını tanımlayabilen NOC (Yeni Nesne Alt Yazılama Modeli, Novel Object Captioner) modeli önerilmiştir. Ünal ve ark. [4] tarafından kaynaklarda ilk kez görüntülerden Türkçe açıklamalar yaratmaya olanak veren ve bu amaçla denektaşı olarak kullanılabilir yeni bir veri kümesi sunulmuştur. TasvirEt adı verilen bu veri kümesi üzerinde, Türkçe görüntü alt yazılama amacıyla kullanılabilir iki yaklaşım önerilmiştir. Kuyu ve ark. [11] Türkçe alt yazı oluşturmak için alt sözcük tabanlı bir model önermişlerdir. Bu makalede modelin eğitim aşamasında, TasvirEt, MS-COCO ve Flickr30k veri kümeleri ile çalışılmıştır. Görüntü alt yazılama için bir özinelemeli sinir ağı olan LSTM modeli kullanılmıştır. Yılmaz ve ark. [12] makalesinde Türkçe alt yazılama için bir model önermişlerdir. Çalışmada MS-COCO veri kümesi Türkçeye çevrilerek kullanılmıştır. Nesne tanıma için CNN ve dil işleme için LSTM modellerinden faydalanılmıştır.

Kaynaklarda bu konuda İngilizcede çok fazla sayıda veri kümesi ve uygulama bulunmaktadır. Türkçede ise çalışmalar yeni başlamış ve gelişmeye açıktır. Yapılmış çalışmalarda, genellikle görüntüdeki nesnelere tasvir edilmektedir. Görüntü içerisinde saptanan nesnelere birbirlerine göre konumları hakkında çok fazla bir bilgi verilmemektedir. Bu nedenle kaynaklarda konum bilgisi içeren çok az sayıda çalışma bulunmaktadır. Çalışmada, kaynaklara yeni bir Türkçe konumlu betimleme veri kümesi kazandırılması amaçlanmıştır. Diğer veri kümelerinden farklı olarak görüntüdeki bazı nesnelere birbirlerine göre konum bilgilerini içermektedir. Makalede Türkçe Konumlu-Betimleme veri kümesi kullanılarak görüntülerin otomatik tasvir edilmesi ve alt yazılama için derin öğrenme yöntemlerinden faydalanılmıştır.

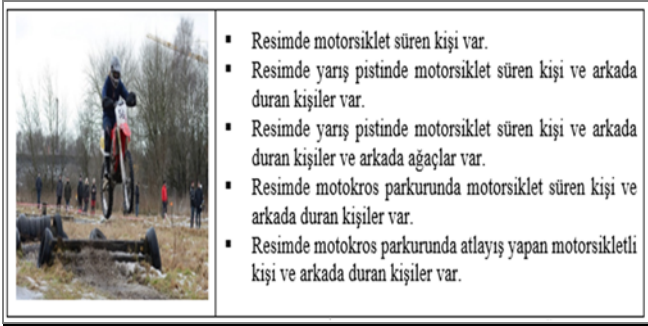
## 3. COCO Türkçe Konumlu-Betimleme Veri Kümesi

Bu bölümde COCO Türkçe Konumlu-Betimleme veri kümesinin nasıl hazırlandığı, kapsamı ve içeriği hakkında ayrıntılı bilgilere yer verilmiştir.

### 3.1. Veri Kümesinin Hazırlanması

Yeni hazırlanan COCO Türkçe Konumlu-Betimleme veri kümesi kısaca Konumlu-Betimleme veri kümesi olarak adlandırılmıştır. Bu çalışma için, MS-COCO [13] veri kümesinden 1500 görüntü seçilmiştir. Veri kümesindeki tüm

görüntüler için, sorunun çözümüne uygun olacak şekilde alt yazılar oluşturulmuştur. Betimleme ve alt yazılama, görüntülerde yer alan bazı nesnelere göre konumlarını içerecek şekilde tanımlanmıştır. Tüm tanımlar kullanıcılar tarafından tek tek yapılmıştır. COCO Konumlu-Betimleme veri kümesi örneği Şekil-1'de gösterilmiştir.



Şekil-1: COCO Türkçe Konumlu-Betimleme veri kümesinden bir örnek görüntü ve alt yazıları.

### 3.2. Oluşturulan Veri Kümesinin Özellikleri

COCO Türkçe Konumlu-Betimleme veri kümesi 1500 görüntü ve bu görüntülere ilişkin 4044 alt yazı içermektedir. Her görüntü, 1-6 arası farklı sayıda alt yazıya sahiptir. Tanım tümceleri en çok 23 sözcükten oluşmaktadır. Tanımlar sade ve kısa tutulmuştur. Nesnelere göre konumları, bakan kişiye göre betimlenmiştir.

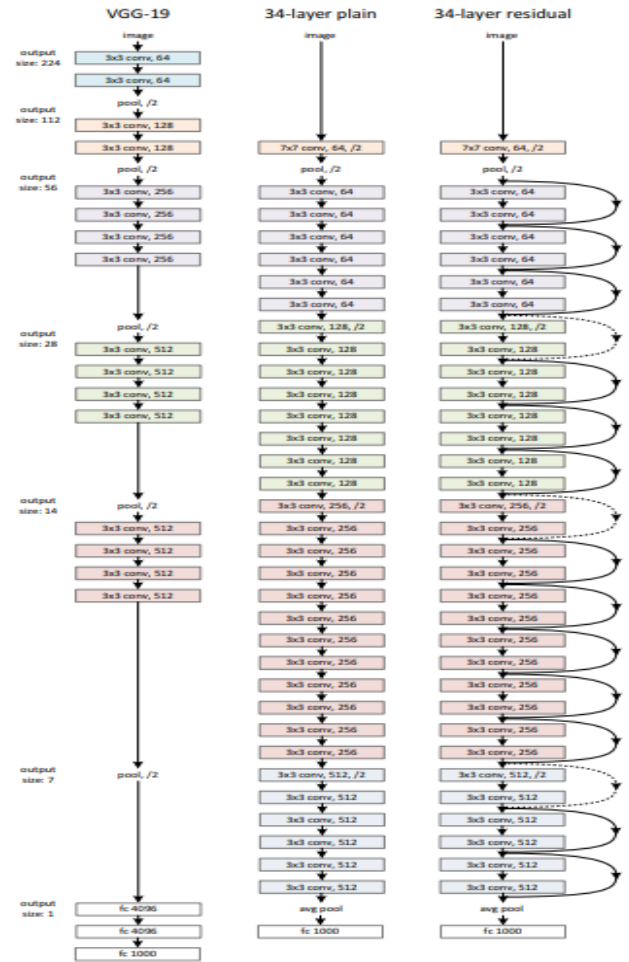
Modelin eğitimi için ilk 1400 görüntü ve sınama için son 100 görüntü kullanılmıştır. Alt yazılardan 3603 tanesi eğitim kümesine kalı, 100 tanesi sınama kümesine aittir. Sınama için seçilen 100 görüntü eğitim kümesinden hariç tutulan görüntülerden ve tüm eğitim kümesini yansıtacak şekilde seçilmiştir. Model eğitimi, her bir resmin tüm alt yazılarının kullanılması amaçlanmıştır. Bundan dolayı her bir alt yazı için görüntü tekrarlı bir şekilde listelenmiştir, böylece eğitim kümesindeki görüntü sayısı 3603'e yükselmiştir.

## 4. Görüntü İçin Otomatik Betimleme ve Alt Yazı Oluşturma Yöntem ve Modelleri

Bu kısımda çalışmada kullanılan yöntemlerden Artık Sinir Ağları (ResNet), Daha Hızlı Bölge-Tabanlı Evrişimli Sinir Ağı, Kapılı Özyinelemeli Birim, önerilen yöntem ile modellerin kapsamı ve içeriği hakkında ayrıntılı bilgilere yer verilmiştir.

### 4.1 Artık Sinir Ağları (ResNet) Modeli

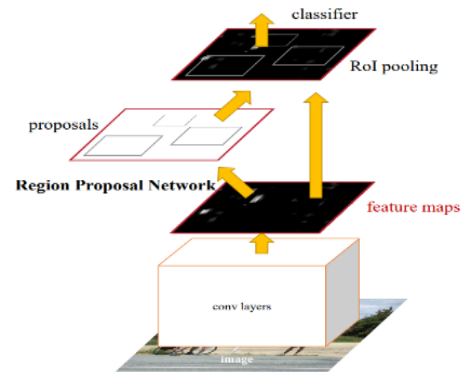
Artık ağlar He, Zhang ve ark. tarafından 2015 yılında geliştirilmiştir. Görsel geometri grubu ağlarına göre daha büyük ve daha derin ancak daha düşük karmaşıklığa sahip olan modeldir. Model katmanlarda referansız öğrenme yerine, artık fonksiyonlardan referans olarak öğrenme sağlanacak şekilde formüle edilmiştir. 50, 101 ve 152 katmana kadar derinliğe sahip versiyonları bulunmaktadır. ResNet101 ve ResNet152 daha fazla üç katmanlı blok kullanılarak oluşturulmuştur. Geliştirilen yöntem sayesinde, artan derinlik nedeniyle oluşan bozulma problemine önemli ölçüde çözüm sağlanmıştır. Artık ağlar mimarisi Şekil-2'de gösterilmiştir [14].



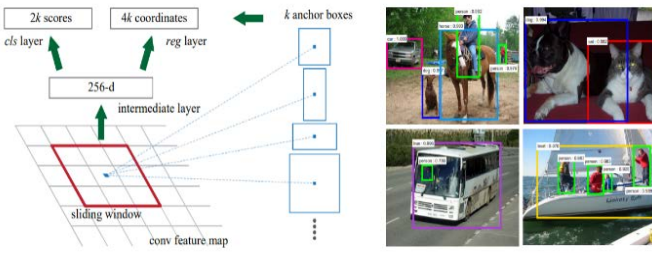
Şekil-2: ResNet Evrişimli Sinir Ağları Mimarisi.

### 4.2 Daha Hızlı Bölge-Tabanlı Evrişimli Sinir Ağları (FasterRCNN) Modeli

Bölge-tabanlı evrişimli sinir ağları, resim üzerindeki farklı nesnelere ve yerlerine tespit etmek için kullanılan modellerdir. FasterRCNN modeli Ren ve ark. tarafından 2015 yılında geliştirilmiştir. FasterRCNN modelinde hız, diğer RCNN modellerine göre daha da artırılmıştır. Bu model bir giriş görüntüsü almakta ve evrişimli sinir ağından geçirilerek bir özellik haritası çıkarmaktadır. Daha sonra ayrı bir bölge önerisi ağı ile ağ bölgeleri belirlenmektedir. FasterRCNN modeli, bölge önerisi ağı ve model çıktısı Şekil 3 ve 4'te gösterilmiştir [15].



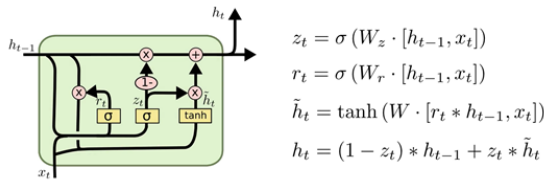
Şekil-3: Faster R-CNN Modeli.



Şekil-4: Faster R-CNN RPN-Bölge Önerisi Ağı ve Model Çıktısı.

### 4.3 Kapalı Özyinelemeli Birim (GRU) Modeli

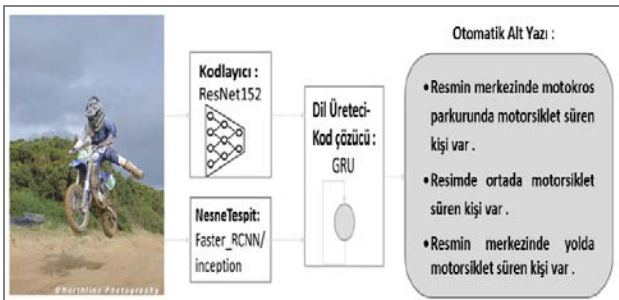
Özyinelemeli sinir ağları döngüsel olarak çalışmaktadır. Ardışık öğrenim yapmak için bu derin öğrenme yapısı kullanılmaktadır. Örneğin, dil öğreniminde cümle içindeki kelimelerin dizilimi önemlidir. Bir özyinelemeli sinir ağı olan GRU modeli, Chung ve ark. tarafından 2014 yılında geliştirilmiştir. Uzun Kısa-Dönemli Bellek [16] modelinin daha özel bir halidir. Klasik RNN modelindeki kaybolan ve patlayan gradyan problemlerine çözüm sağlamıştır. GRU'da sadece gizli durum bulunmaktadır, unutma kapısı ve giriş kapısı birleştirilmiş ve tek bir güncelleme kapısı kullanılmıştır. LSTM modeline göre işlem karmaşıklığı daha az olduğu ve iyi sonuç verdiği için sıklıkla kullanılan bir modeldir. GRU modelinin yapısı Şekil 5'te gösterilmiştir [17, 18].



Şekil-5: GRU Modeli.

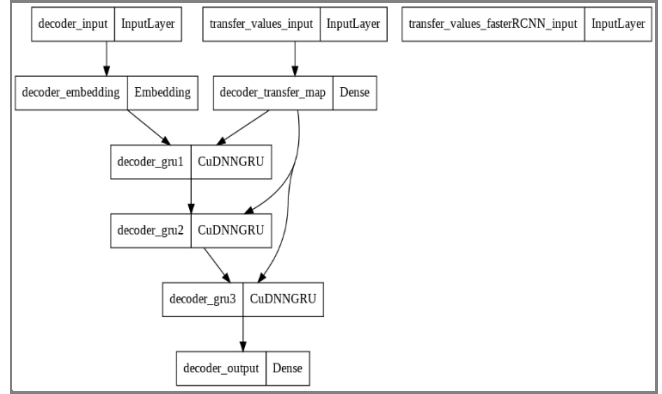
### 4.4 Görüntü İçin Otomatik Türkçe Konumlu-Betimleme ve Alt Yazı Oluşturma Yöntemi

Bu çalışmada görüntünün otomatik Türkçe Konumlu-Betimleme ve alt yazılanması için bir yöntem önerilmiştir. Bu yöntemde kodlayıcı olarak MS-COCO veri kümesi üzerinde eğitilmiş ResNet152 derin öğrenme modeli, doğal dil işleme algoritması olarak üç katmanlı GRU ve nesne tespit için Faster\_RCNN/inception\_resnet\_v2 modeli [19] kullanılmıştır. Yöntem oluşturmada, Pedersen'in "image captioning" [20] modelinden faydalanılmıştır. Çalışmada kodlayıcı olarak kullanılan ResNet152 modelinin son çıktı katmanı iptal edilmiş son ortalama havuzlama (avg\_pool) katmanı başlangıç durumu (initial state) olarak, kod çözücü dil üretici GRU modelinin girdisi olarak alınmıştır. Yöntem model yapısı Şekil-6'da gösterilmiştir.

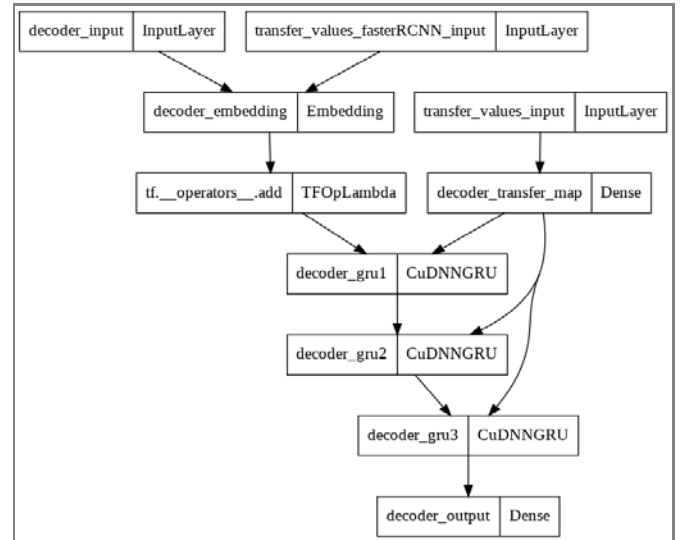


Şekil-6: Otomatik Türkçe Konumlu-Betimleme ve alt yazı oluşturma yöntem mimarisi.

Kelime vektörü oluşturmada, Word2Vec [21, 22] Skipgram algoritması kullanılmıştır. Derlem için TasvirEt [4] veri kümesi, COCO train2017/ val2017 Türkçe alt yazılı (İngilizce- Türkçe makine çevirisi) veri kümesi [23] ve Konumlu-Betimleme veri kümesinin bir kısmı alınmıştır. Derlem toplam 485.918 cümleden oluşmaktadır. Çalışmada kullandığımız yöntem, farklı hiper parametreler [24, 25] ile eğitilmiş ve en iyi sonuç veren 6 model çıktı elde etmek için ortak olarak kullanılmıştır. Yöntem model yapıları Şekil-7 ve 8'de gösterilmiştir.



Şekil-7: Otomatik Türkçe Konumlu-Betimleme ve alt yazı oluşturma modelinin temel yapısı (Model 1-5).



Şekil-8: Otomatik Türkçe Konumlu-Betimleme ve alt yazı oluşturma modelinin temel yapısı (Model 6).

Burada ilk 5 model ortak bir yapıya sahiptir. Sadece farklı olan durum eğitim aşamasında kullanılan hiper parametrelerdir. Bu hiper parametreler Çizelge-1'de verilmiştir.

Çizelge-1: Modellerde Kullanılan Hiper Parametreler

Model No	Kelime Vektörü Değer Aralığı	State	Batch	Epoch	Optimizasyon Öğrenme Oranı
Model1	-1,0:1,0	384	128	49	0,00010
Model2	-1,0: 1,0	384	128	39	0,00012
Model3	-1,0: 1,0	384	128	62	0,00010
Model4	-1,0: 1,0	384	128	65	0,00010
Model5	-0,5: 0,5	384	128	44	0,00012
Model6	-0,1: 0,1	384	128	54	0,00010



## 5. DeneYler

Bu makalede şifreleme ve deşifreleme işlemleri için Matlab 2020b yazılımı kullanılmıştır. Kullanılan bilgisayar Microsoft Windows 10 işletim sistemine, 16 GB belleğe ve Intel(R) Core(TM) i7-4700HQ CPU 2.40 GHz işlemciye sahiptir. Model çıktılarını değerlendirmek için COCO alt yazı değerlendirme kodu ve ölçünleri kullanılmıştır. BLEU ölçüsü, makine çevirisi topluluğundan doğruluk-tabanlıdır. Referanslara göre aday tümce için n-gram tabanlı bir doğruluk hesaplamaktadır [26]. ROUGE-L özetleme topluluğundan, geri çağırma-tabanlı bir otomatik değerlendirme ölçünüdür [27, 28]. METEOR genelleştirilmiş bir bigram eşleştirme kavramına dayanan makine çevirisi değerlendirmesi için otomatik bir ölçüm yöntemidir. Makine çevirisindeki eşleşen sözcüklerin referansa göre ne kadar iyi sıralandığını doğrudan yakalamak için tasarlanmış, bigram-duyarlık, bigram-bulma, eşleştirme kombinasyonu kullanarak bir puan hesaplamaktadır [29]. CIDEr görüntü açıklamalarını değerlendirmek için insan fikir birliğini kullanan, üç ana bölümden oluşan otomatik değerlendirme ölçüsüdür [30]. SPICE (Semantic Propositional Image Caption Evaluation) insan alt yazı değerlendirmelerinin önemli bir bileşenin anlamsal içerik olduğu düşünülerek sahne değerlendirmeleri için önerilen yeni bir alt yazı ölçüsüdür. Bu model çalışmasında verilen isimler ve sahne tanımlarının anlamsal yapılarından yararlanmaktadır [31].

Çalışmada otomatik görüntü alt yazılama modelleri, tüm veri kümesini yansıtacak şekilde seçilen 100 görüntü içeren sinama veri kümesi ve COCO alt yazı değerlendirme dereceleri kullanılarak değerlendirilmiştir. Bunlar BLEU, ROUGE-L, METEOR, CIDEr ve SPICE metrikleridir. Yöntemde modellerin değerlendirme metriklerine göre sonuçları, çıktı elde etmede ortak olarak kullanılan 6'şar modelin ortalaması alınarak elde edilmiştir. 100 adet resim üzerinde gerçekleştirilen sınamalardan elde edilen sonuçlar Çizelge 2 ve 3'te gösterilmiştir.

Çizelge-2: Sinama Veri Kümesi Yöntem Değerlendirme Sonuçları-1

Model No	BLEU1	BLEU2	BLEU3	BLEU4
Model1	0,806	0,678	0,581	0,500
Model2	0,741	0,607	0,507	0,424
Model3	0,710	0,570	0,463	0,383
Model4	0,734	0,605	0,508	0,433
Model5	0,705	0,542	0,427	0,340
Model6	0,678	0,524	0,409	0,319
ORTALAMA	0,729	0,588	0,414	0,400

Çizelge-3: Sinama Veri Kümesi Yöntem Değerlendirme Sonuçları-2


Model No	METEOR	ROUGE_L	CIDEr	SPICE
Model1	0,409	0,792	2,555	0,308
Model2	0,372	0,739	2,263	0,287
Model3	0,372	0,734	2,124	0,287
Model4	0,385	0,751	2,485	0,303
Model5	0,349	0,710	1,972	0,265
Model6	0,352	0,710	1,807	0,258
ORTALAMA	0,373	0,739	2,201	0,285

Otomatik Konumlu-Betimleme ve alt yazı oluşturma amacıyla ortak olarak kullanılan 6 model için, özellikle ROUGE\_L performans değerlendirme ölçüsü ele alındığında ortalama 0,70 üzerinde bir başarımla elde edilmiştir. Ayrıca tüm performans değerlendirme metrik skorlarına göre, önerilen yöntemin başarılı sonuçlar ortaya koyduğu görülmektedir.

Konumlu-Betimleme sinama veri kümesinden seçilen bazı örnekler için önerilen yöntem tarafından üretilen otomatik Türkçe Konumlu-Betimleme ve alt yazı örnekleri ile referans alt yazıları Şekil 9, 10 ve 11'de verilmiştir.

	
<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/000000011631.jpg">http://images.cocodataset.org/train2017/000000011631.jpg</a>	<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/000000395053.jpg">http://images.cocodataset.org/train2017/000000395053.jpg</a>
<ul style="list-style-type: none"> <li>Resimde toprak alanda motosiklet süren kişi var.</li> <li>Resmin merkezinde motokros parkurunda motosiklet süren kişi var.</li> <li>Resimde kırsalda motosiklet süren kişi var.</li> <li>Resimde motokros parkurunda atlayış yapan motosikletli kişi var.</li> <li>Resimde toprak alanda motosiklet süren kişi arkada ağaçlar var.</li> </ul>	<ul style="list-style-type: none"> <li>Resimde masa üstünde kase içinde yatan kedi var.</li> <li>Resimde masa üstünde duran kedi var.</li> <li>Resimde masa üstünde yatan kedi var.</li> <li>Resmin merkezinde masa üstünde yatan kedi var.</li> </ul>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
Model1   Resimde ortada motosiklet süren kişi var . Model2   Resmin merkezinde yolda motosiklet süren kişi var . Model3   Resmin merkezinde motokros parkurunda motosiklet süren kişi var . Model4   Resimde motosiklet süren kişi var . Model5   Resimde ortada motosiklet ortada duran kişi var . Model6   Resimde kırsalda motosiklet süren kişi var .	Model1   Resimde masa üstünde duran kedi var . Model2   Resimde masa üstünde duran kedi var . Model3   Resimde masa üstünde duran kedi var . Model4   Resimde masa üstünde duran kedi var . Model5   Resimde kase içinde duran kedi var . Model6   Resimde masa üzerinde duran kedi var .

Şekil-9: Konumlu-Betimleme sinama kümesi görüntülerinden, referans alt yazıları ile otomatik Konumlu-Betimleme ve alt yazı örnekleri.

	
<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/00000365289.jpg">http://images.cocodataset.org/train2017/00000365289.jpg</a>	<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/00000405276.jpg">http://images.cocodataset.org/train2017/00000405276.jpg</a>
<ul style="list-style-type: none"> <li>• Resimde elinde oyuncak ayıcık olan çocuk var.</li> <li>• Resimde çocuk kucagında oyuncak ayıcık var.</li> <li>• Resimde ortada sandalyede oturan çocuk ve sağ yanında oyuncak ayıcık var.</li> <li>• Resimde elinde oyuncak ayya sarılan çocuk var.</li> <li>• Resimde elinde oyuncak ayıcık olan çocuk arkada duran kişi var.</li> </ul>	<ul style="list-style-type: none"> <li>• Resimde banyoda duran kedi var.</li> <li>• Resimde lavabo üzerinde duran kedi arkada vazo içinde çiçekler var.</li> <li>• Resimde ayna önünde lavabo üzerinde duran kedi önünde vazo içinde çiçekler var.</li> <li>• Resimde ayna önünde lavabo üzerinde duran kedi önünde vazo içinde çiçekler sağda musluk var.</li> </ul>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
<p>Model1   Resimde bir çocuk sağ yanında oyuncak ayıcık var .</p> <p>Model2   Resimde odada duran pasta var .</p> <p>Model3   Resimde koltukta oturan çocuk ve sağ yanında oyuncak ayıcık var .</p> <p>Model4   Resimde elinde oyuncak ayya sarılan bir kız çocuğu var .</p> <p>Model5   Resimde ortada elinde oyuncak ayıcık olan çocuk var .</p> <p>Model6   Resimde solda duran çocuk var .</p>	<p>Model1   Resimde ortada yanyana duran iki kedi var .</p> <p>Model2   Resimde solda lavabo üzerinde duran kedi ve arkada sağda vazo içinde çiçekler var .</p> <p>Model3   Resimde sağda duran kedi ve arkasında duran bisiklet var.</p> <p>Model4   Resimde mutfak tezgahı üzerinde duran kedi ve sağda vazo içinde çiçekler var .</p> <p>Model5   Resimde içeride pencere üzerinde duran kedi ve arkada solda duran kedi var .</p> <p>Model6   Resimde solda yanyana duran iki kedi ve sağda vazo içinde çiçekler var .</p>

**Şekil-10:** Konumlu-Betimleme sınama kümesi görüntülerinden, referans alt yazıları ile otomatik Konumlu-Betimleme ve alt yazı örnekleri.

	
<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/00000129108.jpg">http://images.cocodataset.org/train2017/00000129108.jpg</a>	<b>Referans alt yazıları :</b> <a href="http://images.cocodataset.org/train2017/00000010471.jpg">http://images.cocodataset.org/train2017/00000010471.jpg</a>
<ul style="list-style-type: none"> <li>• Resimde cadde üzerinde bisiklet süren kişi var.</li> <li>• Resimde cadde üzerinde bisiklet süren kadın var.</li> <li>• Resimde cadde üzerinde bisiklet süren kişiler arkada arabalar ve binalar var.</li> <li>• Resimde cadde üzerinde bisiklet süren kadın arkada karşıya geçen kişiler arkada arabalar ve binalar var. Resimde cadde üzerinde bisiklet süren kadın arkada insanlar arabalar ve binalar var.</li> </ul>	<ul style="list-style-type: none"> <li>• Resimde ortada toprak alanda duran filler var.</li> <li>• Resimde ortada toprak alanda duran filler arkada ağaçlar var.</li> <li>• Resimde ortada toprak alanda duran filler arkada dere ve ağaçlar var.</li> <li>• Resimde ortada toprak alanda duran filler önde ve sağda duran kuşlar arkada dere ve ağaçlar var.</li> <li>• Resimde ortada dere kenarında duran filler arkada ağaçlar var.</li> </ul>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
<p>Model1  Resimde cadde üzerinde bisiklet süren kişi arkada arabalar ve arkada binalar var.</p> <p>Model2   Resimde yolda bisiklet süren kişi arkada arabalar ve binalar var.</p> <p>Model3  Resimde yolda bisiklet süren kişi arkada sağda kaldırımda yürüyen insanlar ve arkada ağaçlar var.</p> <p>Model4  Resimde cadde üzerinde bisiklet süren kadın arkada insanlar ve binalar var.</p> <p>Model5   Resimde yolda bisiklet süren kişi arkada insanlar ve binalar var.</p> <p>Model6  Resimde cadde üzerinde bisiklet süren kişi ve arkada arabalar var.</p>	<p>Model1   Resimde ortada hayvanat bahçesinde duran filler arkada duran filler ve arkada ağaçlar var.</p> <p>Model2   Resimde ortada dere kenarında duran fil ve arkada ağaçlar var.</p> <p>Model3   Resimde ortada dere kenarında duran filler arkada ağaçlar var.</p> <p>Model4   Resimde ortada toprak alanda duran filler arkada ağaçlar var.</p> <p>Model5   Resimde ortada dere kenarında duran fil arkada duran fil ve arkada ağaçlar var.</p> <p>Model6   Resimde ortada dere kenarında duran fil arkada duran filler ve arkada ağaçlar var.</p>

**Şekil-11:** Konumlu-Betimleme sınama kümesi görüntülerinden, referans alt yazıları ile otomatik Konumlu-Betimleme ve alt yazı örnekleri.

COCO veri kümesi haricindeki kaynaklardan elde edilen görüntüler üzerindeki otomatik Konumlu-Betimleme ve alt yazı çıktıları Şekil 12 ve 14'te gösterilmiştir.

	
<small>Kaynak : <a href="https://www.sabah.com.tr/fotohaber/yasam/papagan- fiyatları-2023-sultan-jako-sevda-pakistan-forpus-kakadu-macaw-ve-amazon-papaganı-fiyatları/3">https://www.sabah.com.tr/fotohaber/yasam/papagan- fiyatları-2023-sultan-jako-sevda-pakistan-forpus-kakadu-macaw-ve-amazon-papaganı-fiyatları/3</a></small>	<small>Kaynak : <a href="https://www.sunmysports.com/blog/how-to-ride-longboard/">https://www.sunmysports.com/blog/how-to-ride-longboard/</a></small>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
Model1   Resimde dalda duran kuş ve arkada duran insanlar var. Model2   Resimde ağaç üstünde duran kuş var. Model3   Resimde dalda duran kuş var. Model4   Resimde yanyana duran papağanlar var. Model5   Resimde ağaç dalında duran kuş var. Model6   Resimde ağaç dalında duran bir kuş var.	Model1   Resimde beton alanda kaykay süren kişi var. Model2   Resimde elinde tenis raketi olan kişi var. Model3   Resimde beton alanda oturan kişi arkada duran bisiklet var. Model4   Resimde beton alanda kaykay süren kişi arkada sağda bisiklet süren kişi var. Model5   Resimde yolda kaykay süren kişi var. Model6   Resimde beton alanda kaykay süren kişi var.

**Şekil-12:** COCO veri kümesi harici görüntülerde, otomatik Konumlu-Betimleme ve alt yazı örnekleri.

	
<small>Kaynak : <a href="https://www.bigdesign.com/bigdesign-balıklar-mini-semsiye">https://www.bigdesign.com/bigdesign-balıklar-mini-semsiye</a></small>	<small>Kaynak : --</small>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
Model1   Resmin merkezinde kaykay süren kişi ve arkada sağda duran insanlar var. Model2   Resmin merkezinde bir alanda elinde şemsiye olan bir kişi var. Model3   Resmin merkezinde kaykay süren kişi arkada sağda duran insanlar ve binalar var. Model4   Resmin merkezinde elinde tenis raketi olan adam var. Model5   Resmin merkezinde elinde tenis raketi olan adam var. Model6   Resmin merkezinde elinde tenis olan kişi arkada ağaçlar var.	Model1   Resimde ortada yem duran motorsikletler var . Model2   Resimde kaldırımda duran motorsikletler arkada insanlar ve binalar var . Model3   Resimde ortada duran kuşlar ve arkada duran insanlar var. Model4   Resimde deniz üzerinde motorsiklet süren insanlar var. Model5   Resimde kaldırımda duran kuşlar arkada duran insanlar ve binalar var. Model6   Resimde park alanında duran motorsikletler arkada duran insanlar ve arkada kaldırımda duran insanlar var.

**Şekil-13:** COCO veri kümesi harici görüntülerde, otomatik Konumlu-Betimleme ve alt yazı örnekleri.

	
<small>Kaynak : --</small>	<small>Kaynak : --</small>
<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>	<b>Modeller tarafından oluşturulan otomatik alt yazılar :</b>
Model1   Resimde kumsalda oturan kişi arkada duran bisiklet ve arkada deniz var. Model2   Resimde sahilde duran kişi arkada deniz ve deniz var. Model3   Resmin merkezinde bir alanda bankta oturan bir adam ve arkasında deniz var. Model4   Resimde kumsalda oturan adam arkasında deniz ve önünde duran bir kişi var. Model5   Resimde kumsalda yürüyen bir adam arkasında sağda bir bisiklet ve solda bir bina var. Model6   Resimde sahilde duran kadın arkada sağda duran bir kadın ve arkasında deniz var.	Model1   Resimde odada televizyon karşısında duran kişi var. Model2   Resimde odada televizyon karşısında duran kişi var. Model3   Resimde odada televizyon karşısında duran kişi arkada yerde televizyon var. Model4   Resimde odada televizyon karşısında kanepede oturan kişi var. Model5   Resimde odada televizyon karşısında kanepede oturan kişi var. Model6   Resimde odada televizyon karşısında oturan kişi ve arkada masa üzerinde masa üzerinde oturan bir adam var.

**Şekil-14:** COCO veri kümesi harici görüntülerde, otomatik Konumlu-Betimleme ve alt yazı örnekleri.

## 6. Sonuç

Makalede, otomatik Türkçe Konumlu-Betimleme ve alt yazılamaya için yeni Türkçe Konumlu-Betimleme veri kümesi oluşturulmuştur. Hazırlanan veri kümesi otomatik alt yazı oluşturma amacıyla kullanılabilir ve geliştirilebilir. Önerilen yöntem literatürde yer alan diğer çalışmalar ile karşılaştırıldığında yön bilgisi içeren çok az sayıda çalışmadan birisi olma özelliğindedir. Başka bir ifadeyle, Türkçe için yön bilgisini de içeren ilk veri kümesi oluşturulmuş ve önerilen yöntem Türkçe otomatik alt yazılamaya için başarılı bir şekilde kullanılabilir. Veri kümesindeki görüntü çeşitliliği ve alt yazılanmış görüntü sayısı artırıldığında yöntemin başarı oranı daha da artacaktır. Bu çalışma literatüre, Türkçe Konumlu-Betimleme veri kümesi olarak, daha ayrıntılı otomatik tasvir ve alt yazı oluşturma açısından katkı sağlanmıştır. Bu çalışmanın bu alanda çalışma yapacak bilim insanlarına yol göstereceği düşünülmektedir.

## Kaynakça

- [1] Öztemel, E., *Yapay Sinir Ağları*. Papatya Yayıncılık , İstanbul, 2006.
- [2] Nabyev, V., *Yapay Zeka (3. baskı)*. Seçkin Yayıncılık, Ankara, 2010.
- [3] Patterson, J. and A. Gibson, *Deep learning: A practitioner's approach*, " O'Reilly Media, Inc.", 2017.
- [4] Unal, M.E., ve ark., *TasvirEt: Görüntülerden otomatik türkçe açıklama oluşturma için bir denektaçı veri kümesi (TasvirEt: A benchmark dataset for automatic Turkish description generation from images)*. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 2016.



- [5] Antol, S., ve ark., *Vqa: Visual question answering*, in Proceedings of the IEEE international conference on computer vision, 2015.
- [6] Lin, X. and D. Parikh., *Don't just listen, use your imagination: Leveraging visual common sense for non-visual tasks*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
- [7] Vinyals, O., ve ark., *Show and tell: Lessons learned from the 2015 mscoco image captioning challenge*, IEEE, 2016. 39(4): p. 652-663.
- [8] Agrawal, H., ve ark., *Nocaps: Novel object captioning at scale*, in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
- [9] Wu, Y., ve ark., *Decoupled novel object captioner*, in Proceedings of the 26th ACM international conference on Multimedia, 2018.
- [10] Venugopalan, S., ve ark., *Captioning images with diverse objects*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
- [11] Kuyu, M., A. Erdem, and E. Erdem, *Altsözcük Öğeleri ile Türkçe Görüntü Altyazılama Image Captioning in Turkish with Subword Units*, IEEE, 2018.
- [12] Yılmaz, B.D., ve ark., *Image Captioning in Turkish Language*, in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), 2019.
- [13] Lin, T.-Y., ve ark., *Microsoft coco: Common objects in context*, in European conference on computer vision, Springer, 2014.
- [14] He, K., ve ark., *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
- [15] Ren, S., ve ark., *Faster r-cnn: Towards real-time object detection with region proposal networks*, arXiv preprint arXiv:1501.04472, 2015.
- [16] Hochreiter, S. and J. Schmidhuber, *Long short-term memory*, Neural computation, 9(8): p. 1735-1780, 1997.
- [17] Chung, J., ve ark., *Empirical evaluation of gated recurrent neural networks on sequence modeling*, 2014.
- [18] Faulder, D. and T.T. Cladouhos, *Physics-Guided Deep Learning for Prediction of Geothermal Reservoir Performance*, 47 th Workshop on Geothermal Reservoir Engineering Stanford University, 2022.
- [19] The Tensorflow Hub Authors, "[https://tfhub.dev/google/faster\\_rcnn/openimages\\_v4/inception\\_resnet\\_v2/1](https://tfhub.dev/google/faster_rcnn/openimages_v4/inception_resnet_v2/1)", 05.05.2022.
- [20] Pedersen M. E. H., "<https://colab.research.google.com/github/Hvass-Labs/TensorFlowTutorials>", 15.12.2021.
- [21] Mikolov, T., ve ark., *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781, 2013.
- [22] Mikolov, T., ve ark., *Distributed representations of words and phrases and their compositionality*, Advances in neural information processing systems, 2013.
- [23] Samet N., "<https://github.com/nerminsamet>", 05.05.2022.
- [24] Wang, H., K. Ren, and J. Song, *A closer look at batch size in mini-batch training of deep auto-encoders*, in 2017 3rd IEEE international conference on computer and communications (iccc) IEEE, 2017.
- [25] Tieleman, T. and G. Hinton, *Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude*, COURSE: Neural networks for machine learning, 2012.
- [26] Papineni, K., ve ark., *Bleu: a method for automatic evaluation of machine translation*, in Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002.
- [27] Lin, C.-Y. and E. Hovy. *Automatic evaluation of summaries using n-gram co-occurrence statistics*, in Proceedings of the 2003 human language technology conference of the North American chapter of the association for computational linguistics, 2003.
- [28] Lin, C.-Y., *Rouge: A package for automatic evaluation of summaries*, in Text summarization branches out, 2004.
- [29] Banerjee, S. and A. Lavie. *METEOR: An automatic metric for MT evaluation with improved correlation with human judgments*, in Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, 2005.
- [30] Vedantam, R., C. Lawrence Zitnick, and D. Parikh. *Cider: Consensus-based image description evaluation*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
- [31] Anderson, P., ve ark., *Spice: Semantic propositional image caption evaluation*, in European conference on computer vision Springer, 2016.