



ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: <https://dergipark.org.tr/ij3dptdi>

ÜRÜN GÖRSELLERİNİ KULLANARAK E-TİCARET SİSTEMLERİ İÇİN ÜRÜN BAŞLIĞI OLUŞTURULMASI

CREATING PRODUCT TITLES FOR E-COMMERCE SYSTEMS FROM IMAGES

Yazarlar (Authors): Caner Balım^{ID*}, Kemal Özkan^{ID}

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): Balım C., Özkan K., "Ürün Görsellerini Kullanarak E-Ticaret Sistemleri İçin Ürün Başlığı Oluşturulması" *Int. J. of 3D Printing Tech. Dig. Ind.*, 5(3): 614-624, (2021).

DOI: 10.46519/ij3dptdi.991789

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): <https://dergipark.org.tr/en/pub/ij3dptdi/archive>

ÜRÜN GÖRSELLERİNİ KULLANARAK E-TİCARET SİSTEMLERİ İÇİN ÜRÜN BAŞLIĞI OLUŞTURULMASI

Caner Balm^a , Kemal Özkan^b 

^a Afyon Kocatepe Üniversitesi, Sandıklı MYO, Bilgisayar Teknolojileri Bölümü, TÜRKİYE

^b Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, TÜRKİYE

* Sorumlu Yazar: cbalim@aku.edu.tr

(Geliş/Received: 06.09.2021; Düzeltme/Revised: 02.12.2021; Kabul/Accepted: 24.12.2021)

ÖZ

E-ticaret sistemlerindeki ürünlerde bilgilerin zaman zaman insan kaynaklı hatalarla karşılaşılmaktadır. Ürün başlığının yanlış girilmesi, fiyatın düşük girilmesi vb. problemler hem kullanıcıları hem de satıcıları olumsuz yönde etkilemektedir. Bu makalede insan kaynaklı yanlışlıkların önüne geçilmesi amacıyla e-ticaret sistemlerinde eklenen ürünlerinin başlıklarının otomatik oluşturulmasına yönelik bir yöntem sunulmuştur. Son zamanlarda özellikle kodlayıcı- kod çözücü mimarilerin başarılı sonuçlar vermesiyle araştırmacılar tarafından ilgi gören görüntü altyazısı oluşturma sistemleri otonom arabalar ve görme engellilere yardım konuları dahil birçok alanda kullanılmaktadır. Çalışmada otomatik ürün başlığı oluşturulmasının yanı sıra sisteme eklenen özellikler ile ürün görsellerinin metinsel anlatım başarısının ne ölçüde etkileneceği konusu üzerine durulmuştur. Önerilen sistemin başlık oluşturma performansı BLEU, METEOR, ROUGE ve CIDEr gibi bu alanda kullanılan en yaygın değerlendirme ölçütleri kullanılarak değerlendirilmiştir. Bu sistemlerin e-ticaret sitelerindeki içerik iş yükünü azaltacağı düşünülmektedir.

Anahtar Kelimeler: Görüntü Altyazı Üretme, Bilgisayarla Görü, Doğal Dil İşleme, Özyinelemeli Sinir Ağları, Evrişimsel Sinir Ağları.

CREATING PRODUCT TITLES FOR E-COMMERCE SYSTEMS FROM IMAGES

ABSTRACT

In e-commerce systems, human-induced errors are often encountered in the product information. Users and sellers are negatively affected by problems such as incorrect entry of product title or low price. In this article, an automatic captioning system has been proposed about product titles in e-commerce systems in order to prevent human-induced mistakes. Recently, especially with the successful results of encoder-decoder architectures, image captioning systems are used in many areas such as autonomous cars and helping the visually impaired. In addition to the creation of product titles within the scope of the study, the issue of the extent to which the textual depiction success of the product images with the features added will be affected are emphasized. The performance of the proposed system was evaluated using the most common evaluation metrics such as BLEU, METEOR, ROUGE and CIDEr. It is thought that these systems can reduce the content workload on e-commerce sites.

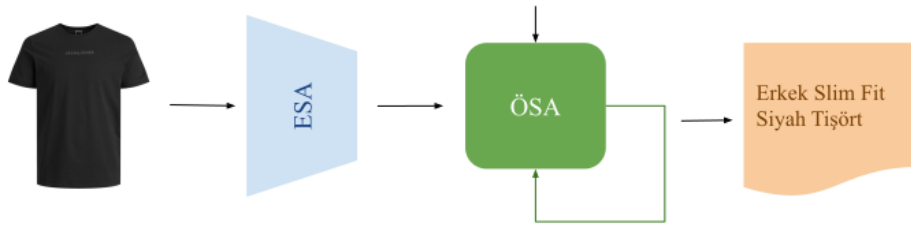
Keywords: Image Captioning, Computer Vision, Natural Language Processing, Convolutional Neural Network, Recurrent Neural Networks.

1. GİRİŞ

Bilgisayar sistemlerin gelişmesiyle birlikte yapay zeka kullanan sistemler gün geçtikçe hayatımızda daha fazla yer edinmektedir. Başta sağlık, tarım gibi hayati derecede öneme sahip alanların yanında elektronik ticaret üzerine yapılmış onlarca çalışma bulunmaktadır [1,2]. E-ticaret sistemlerinin hızla

gelişmesi ve yaygınlaşması ile büyük miktarda veri internet üzerinden dünya ile paylaşılmaktadır. Bilim insanları bu verileri görüntü işleme, doğal dil işleme ve sınıflandırma gibi makine öğrenme çalışmaları için kullanmaktadır. Görüntü altyazısı oluşturma (Image Captioning) Doğal Dil İşleme ve Bilgisayarlı Görme teknikleri kullanarak bir görüntünün metinsel tanımını oluşturma işlemidir. İnsanların dünyayı nasıl gördüklerini makineye öğretmek olan bu sistemlerin amaçlayan bu sistemlerin en önemli örneklerinde görme engelli insanlara yardım ve insan robot etkileşimi gibi konular yer almaktadır [3, 4].

Görüntü altyazısı oluşturma geçmişte çalışılan bir konu olsa da derin öğrenme teknikleri ile elde edilen başarılı sonuçlar bu alandaki ilgiyi arttırmıştır. Başarılı sonuçlar veren çalışmaların derin öğrenme tabanlı makine çevirisi sistemlerinin görüntü altyazısı oluşturma sistemlerine uyarlanması ile gerçekleştiği görülmüştür [5, 6]. Bu yöntemlerde Evrişimsel Sinir Ağları (ESA) ile çıkarılan görüntü özneliklerin makine çevirisinde problemlerinde başarılı sonuçlar üreten Kodlayıcı-Kod Çözücü modellere girdi olarak verildiği ve sıradan sıraya problemler olarak ele alındığı görülmüştür. Örnek çıktı Şekil 1’ de gösterilmiştir. Son zamanlarda üretilen çalışmalarda ise dikkat mekanizmasını da bu sistemlere adapte edildiği ve çok daha iyi sonuçlar elde edildiği görülmüştür [7, 8].



Şekil 1. Önerilen sistem temel şeması

Bu çalışmada ürün görüntüleri ve özellikleri kullanılarak otomatik ürün başlığı oluşturulmasına yönelik bir çalışma gerçekleştirilmiştir. Çoğunluğu moda ürünlerinden oluşan veri seti üzerinde yapılan çalışmalarda başlangıçta ürün görüntülerinden ESA ağları kullanılarak anlamsal öznelik vektörlerinin çıkarılması gerçekleştirilmiştir. Çalışmanın sonraki aşamasında ürün başlığını oluşturan kelimelerden oluşan metin vektörleri ve görüntü öznelik vektörleri Özyinelemeli Sinir Ağı (ÖSA) türü olan Uzun Kısa Süreli Yinelemeli Sinir Ağı (UKSB) kullanılarak eşleştirilmiştir. Veri setinde yer alan marka adı, rengi gibi özelliklerin altyazı oluşturma performansına etkisini ölçmek amacıyla çeşitli deneyler gerçekleştirilmiştir. Buna ek olarak dikkat mekanizması ve önceden eğitilmiş kelime gömme vektörleri de modele dahil edilerek kapsamlı deneyler gerçekleştirilmiştir.

Çalışmadaki motivasyon aşağıda maddeler halinde sunulmuştur:

- E-ticaret sitelerindeki ürünlerin başlıklarının ürün görüntüleri üzerinden otomatik oluşturulmasının sağlanıp sağlanamayacağını ölçülmesi,
- Görüntülerdeki marka izlerinin görüntü altyazısı oluşturma sistemleri ile saptanma performansının ölçülmesi,
- Görüntü altyazısı oluşturma sistemlerine metin ve görüntü dışında verilen özelliklerin sistemin performansı üzerindeki etkisinin araştırılması.

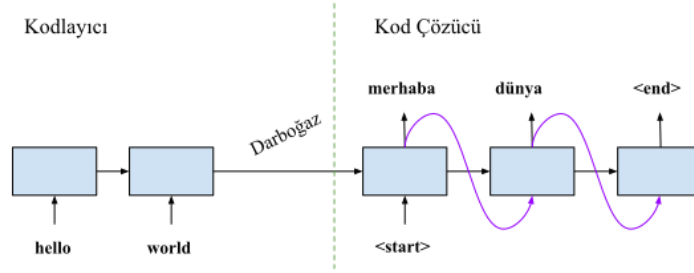
Takip eden Bölüm 2’de literatürde yer alan görüntü altyazısı oluşturma çalışmalarına yer verilmiştir. Bölüm 3’de kullanılan materyal ve metot, Bölüm 4’de yapılan deneyler ve değerlendirmelerden bahsedilmiştir. Bölüm 5’de ise sonuçlar üzerine durulmuştur.

2. LİTERATÜR TARAMASI

Literatür incelendiğinde bu alandaki ilk çalışmaların direk olarak insan açıklamaları ile yapıldığı, ilk makine tabanlı uygulamaların ise kural tabanlı sistemlerle gerçekleştirildiği görülmüştür. Zamanla nesne tespiti ve doğal dil işleme alanlarındaki gelişmelerle birlikte görüntü altyazısı oluşturma sistemlerinin başarısı yükselmiştir. Özellikle görüntülerde tüm nesnelerin tanınması ve bu tanımlamaların birleştirilip cümle haline getirilmesi çalışmaları bu alanda büyük ilerlemelerin gerçekleşmesini sağlamıştır. Farhadi vd. [9] çalışmalarında görüntüleri nesne, aksiyon ve sahne üçlüsüne ayırmış, bunların cümle uyumluluğunu hesaplayan bir skor sistemi geliştirmiştir. Feng ve Lapata [10] çalışmalarında haber

görüntüleri için yeni otomatik altyazı oluşturma işlemi için çeşitli içerik belirleme ve yüzey çıkarma yöntemlerini birlikte kullanmışlardır. 2010 yılından sonra sinir ağları alanındaki ilerlemelerle birlikte nesne algılama, nesne tanıma ve cümle oluşturma konularında ilerlemeler kaydedilmiş, bu gelişmeler görüntü altyazısı oluşturma sistemlerin başarısını yükseltmiştir. Derin öğrenme çalışmalarının ise temelde benzer makine çevirisinde kullanılan Kodlayıcı-Kod Çözücü modeli mimarisinden esinlendiği görülmüştür [5,9,10].

Kodlayıcı-Kod Çözücü modelleri sıradan sıraya (Seq2Seq) problemleri çözmek için oluşturulmuştur. Sıradan Sıraya problemler, girdi ve çıktının bir dizi olduğu özel bir modelleme problemi sınıfıdır [11]. Bu yöntemde kodlayıcı girdi olarak aldığı cümle vektörünü sıkıştırır. Kod çözücü ise Şekil 2' de görüldüğü üzere tahmin edilen girdiden kodlayıcıya göre çıktıyı tahmin etmeye çalışır.



Şekil 2. Sıradan sıraya problemler

Vinyals vd. [11] görüntü özelliklerini çıkarmak için ESA mimarisi üzerine kurulu bir kodlayıcı ve görüntü altyazılarını oluşturmak için bir kod çözücü olarak UKSB kullanan ilk çalışma olması açısından önem taşımaktadır. Sönmez ve ark. [12] görüntü altyazısı oluşturma çalışmalarında sıkça kullanılan MS COCO veri setini Türkçeleştirmiş, UKSB kullanarak Türkçe görüntü altyazısı oluşturma modeli geliştirmiştir.

Son yaklaşımlarda görüntü altyazısı oluşturmak için dikkat mekanizması ve Kodlayıcı-Kod Çözücü kombinasyonunun kullanıldığı görülmüştür [7, 14, 15]. Moda alanında yapılan görüntü altyazısı oluşturma çalışmalarına da rastlanılmıştır. Yang vd. [16] giysiler için altyazı oluşturma performansını iyileştirmek için bir dizi strateji önerilmiş ve moda altyazı oluşturma çalışmalarında kullanılacak bir veri kümesi oluşturmuştur. Li vd. [17] ise giysiler hakkında altyazı oluşturmak için görsel öznitelikleri AttrNET adlı özel bir yapıda saptayan ve LSTM + dikkat mekanizması kullanan bir model önermiştir [18, 19].

3. MATERYAL VE METOT

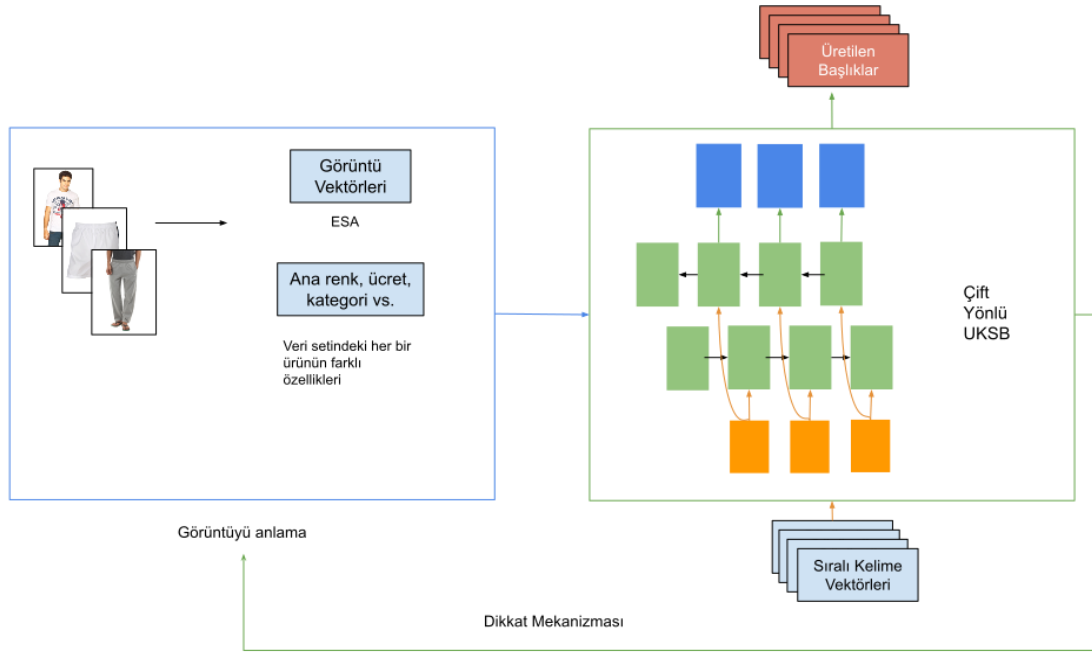
Bu kısımda öncelikle sunulan yöntemlerin uygulandığı materyal açıklanacak sonrasında yöntemin genel basamakları ayrıntılı şekilde anlatılacaktır.

3.1. Materyal

Bu çalışma için Fashion Product Images Dataset (FPID) kullanılmıştır. Veri seti içerisinde 44 000 adet ürün bulunmakta olup her ürün görüntü, başlığı, 8 farklı ana kategoriden biri, 45 farklı alt kategoriden biri, sezon ve yıl gibi bir moda ticaret sitesinde olabilecek özellikleri içermektedir.

3.2. Metot

Önerilen çalışmada temelde Kodlayıcı-Kod Çözücü modeli kullanılmış olup kodlayıcı kısmında ESA ve kod çözücü kısmında ise UKSB + dikkat mekanizması kullanılmıştır. Başlangıçta ürün başlıkları metin ön işleme aşamasından geçirilmiş (1), ürün görüntüleri popüler ESA yöntemlerinin ImageNET veri setindeki ağırlıkları kullanılarak vektör dizisi haline dönüştürülmüş (2), ardından elde edilen veriler kod çözücü kısmında çift yönlü UKSB ve dikkat mekanizması kullanılarak çözülmüştür (3). Önerilen yöntemin şeması Şekil 3'de gösterilmiştir.



Şekil 3. Önerilen model şeması

3.2.1. Metin Ön İşleme

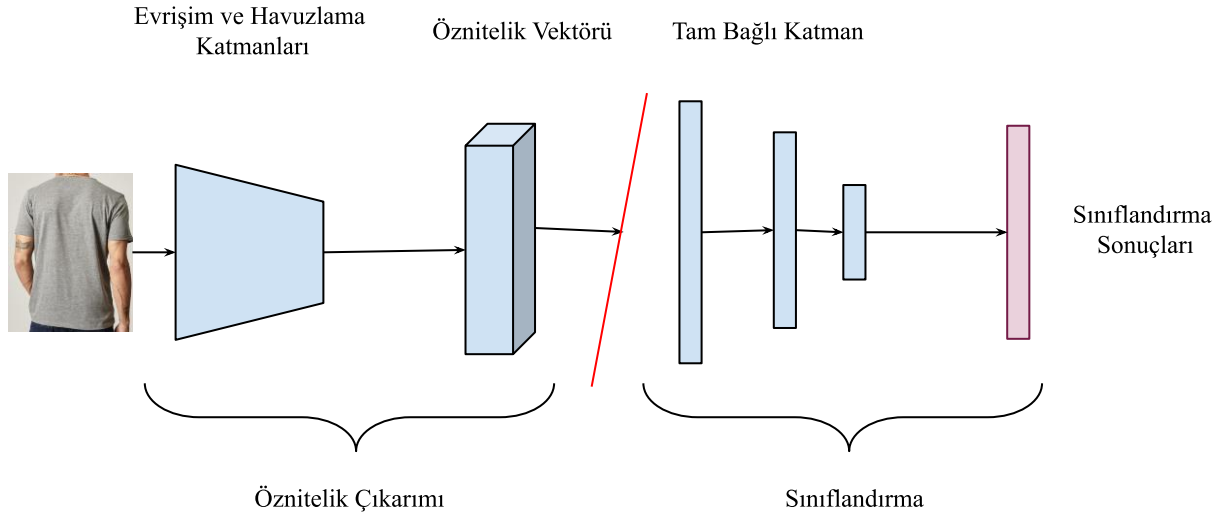
Metinlerin makine öğrenme problemlerine dahil edilmesi için bazı ön işlemlerden geçmesi gerekmektedir. Önerilen çalışmada veri setindeki metin tanımlamaları aşağıdaki işlemlerden geçirilmiştir:

1. Tüm kelimeler küçük harfli hale getirilmiştir.
2. Tüm noktalama işaretleri çıkarılmıştır.
3. Tüm rakamlar çıkarılmıştır.
4. Her bir ürün başlığının başına <start> ve sonuna <end> etiketleri eklenmiştir. <start> etiketi dizinin başlangıcını belirtirken <end> ise dizinin sonlandığını göstermektedir.

3.2.2. Görüntü Özniteliklerin Çıkarılması

Sinir ağları ise hangi özelliklerin diğerlerinden önemli olduğunu öğrenme yeteneğine sahiptir. İnsan çabası gerektirecek karmaşık görevler için bile, bir öğrenme algoritması kısa sürede iyi bir özellik kombinasyonu bulabilir. Daha önce farklı görevlerde başarıları kanıtlanmış yapay sinir ağlarının ağırlıkları farklı bir problemde kullanılabilir. Literatürde sıklıkla kullanılan bu yöntem transfer öğrenme adı verilmektedir. Önerilen çalışmada görüntü temsil vektörlerinin çıkarılması için literatürde sıkça tercih edilen transfer öğrenme tekniği kullanılmıştır. Transfer öğrenme tekniği Şekil 4' de gösterilmiştir.

Bu çalışmada transfer öğrenme tekniğinden ImageNet veri setinde başarıları kanıtlanmış VGG16 ve InceptionV3 ESA modelleri kullanılarak yararlanılmıştır [12, 20]. Bu modellerin öznitelik çıkarımı için kullanılan evrimsel katmanları ile sınıflandırma için kullanılan tam bağlı katman arasındaki bağlantı kesilmiş ve ImageNet ağırlıkları kullanılarak yüksek seviyeli öznitelikler elde edilmiştir. Sonuçta öznitelik çıkarılması işlemi sonucunda veri setindeki her bir ürün görüntüsü için VGG16' dan [1,4096] ve InceptionV3' den ise [1,2048] boyutunda görüntü temsil vektörleri elde edilmiştir.



Şekil 4. Görüntü özniteliklerinin çıkarılması için yapay sinir ağları ve transfer öğrenme tekniğinin kullanılması

3.2.3. Kod Çözücünün Tanımlanması

Önerilen çalışmada kod çözücü olarak ÖSA' da sıklıkla karşılaşılan gradyanların yok olma problemi sahip olduğu kapılarla çözen UKSB ağları kullanılmıştır. UKSB' yi kullanan Kodlayıcı-Kod Çözücü yaklaşımındaki temel sorun tüm bilginin sabit uzunluktaki bir bağlam vektörüne sıkıştırılmasının gerekliliğidir. Bu durum özellikle uzun cümleler oluşturulurken kötü sonuçlara yol açabilmektedir.

Bu sorunu çözmek için son yıllarda dikkat (attention) mekanizması kullanılmaktadır. Bu çalışmada Bahdanau dikkat mekanizması kullanılmıştır [21]. Bahdanau vd. bir sonraki adımın gizli durumu (h_t)' yi hesaplamak için (h_{t-1})' deki gizli durumunu ve bağlam vektörünü (c_t) aşağıda formülü verilen şekilde kullanmıştır.

$$h_t = LSTM(h_{t-1}, [c_t; h_t]) \quad (1)$$

Denklem 2' de dikkat ağırlıklarının hesaplanması ($\alpha_{t,i}$) ve gizli durumlar arasındaki ilişkiyi kuran (*align*) fonksiyonu gösterilmiştir. Denklem 3' de ise bağlam vektörünün (c^t) hesaplanması gösterilmiştir. Bu denklemlerden bağlam vektörünün (c^t) gizli durumların ($\alpha_{t,i}$) ağırlık ortalamasından hesaplandığı görülmektedir.

$$\alpha_{t,i} = align(h_t, h_s) = \frac{e^{a(h_t, h_s)}}{\sum_{s'} e^{a(h_t, h_{s'})}} \quad (2)$$

$$c_t = \sum_{i=1}^N \alpha_{t,i} h_i \quad (3)$$

Ürün başlıklarının oluşturulması için ön işlemlerden geçen her bir ürün başlığı kelimelerden oluşan bir vektör dizisine dönüştürülmüştür. Sonrasında bu vektörlerdeki her bir benzersiz kelime sıralı tam sayılara dönüştürülmüştür. Kelime vektörlerin sayılarını eşitlemek için vektörlere dolgulama(padding) işlemi uygulanarak sıfır eklenmiştir. Ardından resimden çıkarılan öznitelikler ve kelime vektörleri sıradan sıraya modele girdi olarak verilmiştir. Ek olarak değerlendirme aşamasında kelimeler arasındaki mantıksal ilişkilerin daha iyi ortaya konulmasına yardımcı olan, Stanford Üniversitesi tarafından geliştirilen açık kaynak kodlu Global Word for Vector Representations(GLOVE) modelleri de kullanılmıştır[22].

Önerilen modelde hem UKSB hem de UKSB + dikkat mekanizması ile deneyler gerçekleştirilmiştir. Ayrıca önerilen modelde tek yönlü UKSB yerine Çift Yönlü UKSB tercih edilmiştir. Çift Yönlü UKSB' de öğrenme orijinal verilerle bir kez baştan sona ve bir kez baştan sona beslenmekte olup bu sayede eğitim işlemi daha kısa sürmekte ve daha iyi bir öğrenme gerçekleştirilebilmektedir. Ek olarak veri

setinde bulunan marka, renk, kategori, ücret gibi özniteliklerin gerekli ön işlemlerden geçirilip modele dahil edilmesi ile sonuçların başarı oranının ne ölçüde değiştiği ölçülmüştür.

3.2.4. Değerlendirme Metrikleri

Basit doğal dil işleme problemlerini değerlendirmek sadece etiket eşleştirilmesi yapıldığı için kolaydır. Fakat otomatik görüntü altyazısı oluşturma çok daha kompleks bir problemdir ve değerlendirmesi daha karmaşık tekniklerle gerçekleştirilmektedir. Görüntü altyazısı oluşturma modellerinin değerlendirilmesi için kullanılan yaygın şekilde kullanılan metrikler BLUE, ROUGE, METEOR ve CIDEr metrikleridir [23-26]. Bu otomatik metriklerin olası skorları 0 ile 1 arasında değişmektedir. Üretilen çeviriler ile insan çevirisinin birbirlerine ne kadar örtüştüğü üzerinden yapılan hesaplamada tam örtüşmeye 1.0 skor verilirken, tam uyumsuzluğa ise 0.0 puan verilmektedir.

BLEU bu teknikler arasında en çok tercih edilen teknik olarak öne çıkmaktadır. Modifiye edilmiş n-gram hassasiyeti (precision) üzerinden hesaplanmaktadır. Ayrıca en uzun eşleşmeyi doğru bir şekilde hesaplamak için kısalık (brevity) cezasını kullanmaktadır. ROUGE geri çağırma (recall) tabanlı bir metrik olup genellikle metin özetleme çalışmalarını değerlendirmede kullanılmaktadır. Bu çalışmada en uzun doğru eşleşmeyi sapmamak için ROUGE metriğinin ROUGE-L türevi kullanılmıştır. METEOR, eşleme unigramlarına dayalı ağırlıklı bir F-skoru ve yanlış kelime sırası için bir ceza işlevi ile kullanarak değerlendirmeleri gerçekleştirmektedir. CIDEr metriği ise bu metrikler arasında en son üretilen metrik olup özellikle görüntü altyazı üretme sistemlerinin değerlendirilmesi için üretilmiştir. CIDEr, aday ve referans cümleleri arasındaki benzerliği ölçmek ve her n-gram için istatistikleri toplamak için Terim Frekansı-Ters Belge Frekansı (TF-IDF) metriğini kullanmaktadır.

4. DENEYLER

İlk önce deneylerin gerçekleşmesi için çalışmanın sınındığı FPID üzerinde bozuk ve eksik verilerin düzenlenmesi işlemi gerçekleştirilmiştir. Bu işlemde sonra veri setinde bulunan ürünlerin eğitim, test ve validasyon dağılım sayıları Çizelge 1’de gösterilmiştir.

Çizelge 1. Çalışmada kullanılan verilere ait bilgiler

	Eğitim	Test	Validasyon	Toplam
Görüntü Sayısı	28,855	8,487	5,093	42,435

Deneyisel çalışmalar, NVIDIA Quadro P5000 üzerinde Google tarafından geliştirilen ve tensör işlemlerini hızlı bir şekilde gerçekleştirebilen TensorFlow ve yüksek seviyeli bir derin öğrenme aracı olan Keras ile gerçekleştirilmiştir. Veri setindeki başlıkların tümünde yapılan ön işleme aşamalarından sonra 5856 benzersiz kelime olduğu ve en uzun başlıktaki kelime sayısının 14 olduğu saptanmıştır.

Deneylerde UKSB modelinde farklı sıkıştırma boyutlarının denenmesinin yanı sıra popüler bir yaklaşım olan büyük veri kümelerinde eğitilmiş kelime ağırlık vektörleri GLOVE100 ve GLOVE300 kullanılmıştır. Model eğitilirken yapılan kontroller neticesinde epok sayısı 40, optimizasyon algoritması adam ve öğrenme oranı ise 0.0001 olduğu durumlarda sistemin daha performanslı çalıştığı gözlemlenmiştir. Ayrıca aşırı öğrenmeyi önlemek için kullanılan bir düzenleme yöntemi olan bırakma (Dropout) 0.2 ve 0.4 oranlarında kullanılmıştır. Deneylerin tamamında kod çözücü kısmında açgözlü yaklaşım yerine ışın arama (beam search) yaklaşımı ışın genişliği (beam width) değeri 3 alınarak uygulanmıştır. Deneylerin sonuçlarının gösteriminde BLEU-1, BLEU-2, BLEU-3, BLEU-4, METEOR, ROUGE-L ve CIDEr metrikleri çizelgelerde sırasıyla B-1, B-2, B-3, B-4, M, R ve C olarak kısaltılmıştır. Elde edilen en başarılı sonuçlar çizelgelerde kalın punto ile gösterilmiştir.

4.1. Ürün Görsellerinden Otomatik Başlık Oluşturulması

Bu deneyde girdi olarak sadece ürün görsellerini alan ve çıktı olarak ürün başlıkları üreten Kodlayıcı-Kod Çözücü model test edilmiştir. Deneyde kodlayıcı ESA modeli, UKSB gömme boyutu, dikkat mekanizması ve GLOVE vektörlerinin çeşitli kombinasyonu gibi farklı hiper parametreler test edilmiş olup sonuçlar Çizelge 2’ de gösterilmiştir.

Çizelge 2. Ürün görsellerinden ürün başlığı oluşturma deneyi değerlendirme sonuçları

Metot	Gömme Boyutu	B-1	B-2	B-3	B-4	M	R	C
Inceptionv3 + UKSB	100	0.4978	0.3503	0.2599	0.1998	0.2419	0.5053	0.1872
VGG16 + UKSB	100	0.5250	0.3776	0.2825	0.2196	0.2560	0.5320	0.2068
Inceptionv3 + UKSB+GLOVE	100	0.5021	0.3558	0.2646	0.2048	0.2449	0.5126	0.1924
VGG16 + UKSB+ GLOVE	100	0.5163	0.3709	0.2776	0.2154	0.2526	0.5293	0.2054
Inceptionv3 + UKSB	300	0.4978	0.3503	0.2599	0.1998	0.2419	0.5053	0.1872
VGG16 + UKSB	300	0.4713	0.3229	0.2353	0.1782	0.2317	0.4791	0.1675
Inceptionv3 + UKSB+GLOVE	300	0.5135	0.3670	0.2751	0.2130	0.2499	0.5200	0.2007
VGG16 + UKSB+ GLOVE	300	0.4966	0.3472	0.2552	0.1952	0.2406	0.5019	0.1825
Inceptionv3 + UKSB +Dikkat	100	0.5123	0.3627	0.2686	0.2060	0.2487	0.5182	0.1952
VGG16 + UKSB + Dikkat	100	0.4970	0.3487	0.2585	0.1993	0.2417	0.5054	0.1902
Inceptionv3 + UKSB+ Dikkat+ GLOVE	100	0.5318	0.3887	0.2949	0.2319	0.2614	0.5436	0.2204
VGG16 + UKSB + Dikkat +GLOVE	100	0.5223	0.3772	0.2839	0.2228	0.2552	0.5305	0.2104
Inceptionv3 + UKSB +Dikkat	300	0.5092	0.3635	0.2727	0.2129	0.2503	0.5243	0.2035
VGG16 + UKSB + Dikkat	300	0.5184	0.3696	0.2745	0.2110	0.2518	0.5258	0.2005
Inceptionv3 + UKSB+ Dikkat+ GLOVE	300	0.5309	0.3831	0.2877	0.2240	0.2591	0.5380	0.2106
VGG16 + UKSB + Dikkat +GLOVE	300	0.5210	0.3743	0.2804	0.2174	0.2543	0.5291	0.2057

Önerilen modelin kodlayıcı olarak InceptionV3' ün kullandığı durumlarda VGG16' nın kullanıldığı durumlara göre daha başarılı sonuçlar ürettiği gözlemlenmiştir. Ayrıca gömme boyutu 100 olan GLOVE vektörü ile dikkat mekanizmasının birlikte kullandığı durumda model başarısının daha yüksek olduğu görülmüştür.

4.2. Ürün Görselleri ve Özelliklerinin Kodlayıcı Olarak Kullanılmasının Altyazı Oluşturma Performansı Üzerine Etkisinin İncelenmesi

FPID içinde ürün başlıkları dışında marka, renk, mevsim gibi her moda ürününde olabilecek özelliklerin olduğu saptanmıştır. Bu özelliklerin kodlayıcı kısmına görüntü öznitelikleri ile birlikte verilmesinin altyazı oluşturma performansını hangi ölçüde değiştireceği ölçülmek istenmiştir. FPID yer alan metin dışındaki 11 özellik ve ayrıntıları Çizelge 3' de gösterilmiştir.

Çizelge 3. Veri setindeki metin dışındaki özelliklerin detayları

Özellik	Sayısı
Kullanıcı Cinsi	5
Ana Kategori	6
Alt Kategori	44
Eşya Türü	139
Temel Renk	46
Sezon	4
Yıl	13
Kullanım Tipi	7
Marka Adı	415
Yaş Grubu	6
Ücreti	-

Çizelgede gösterilen özellikler modele verilmeden önce türlerine göre gerekli etiketleme ve normalizasyon işlemlerinden geçirilmiştir. Ardından önce hepsi, sonrasında tek tek modele dahil edilerek performans değerlendirilmesi gerçekleştirilmiştir. Çizelge 4’ de belirtilen özelliklerinin tümünün modele verilmesi ile elde edilen sonuçlar gösterilmiş, Çizelge 5’ de ise bu 11 özelliğin kodlayıcı olarak InceptionV3 ile kod çözücü olarak gömme boyutu 300 olan GLOVE ve UKSB + dikkat mekanizması kullanılarak tek tek modele verilmesi sonucunda her bir özelliğin sonuca etkisi gösterilmiştir.

Çizelge 4. Ürün görselleri ve tüm özelliklerinin kodlayıcı olarak kullanılması neticesinde altyazı oluşturma performansının değerlendirmesi

Metot	Gömme Boyutu	B-1	B-2	B-3	B-4	M	R	C
Inceptionv3 + UKSB	100	0.7981	0.6960	0.5995	0.5276	0.4346	0.8264	0.5245
VGG16 + UKSB	100	0.7890	0.6862	0.5901	0.5187	0.4290	0.8196	0.5154
Inceptionv3 + UKSB+GLOVE	100	0.8069	0.6963	0.5919	0.5147	0.4322	0.8188	0.5047
VGG16 + UKSB+ GLOVE	100	0.7936	0.6826	0.5788	0.5018	0.4245	0.8114	0.4943
Inceptionv3 + UKSB	300	0.7954	0.6912	0.5943	0.5224	0.4317	0.8205	0.5174
VGG16 + UKSB	300	0.7880	0.6806	0.5829	0.5118	0.4250	0.8104	0.5044
Inceptionv3 + UKSB+GLOVE	300	0.8038	0.6938	0.5922	0.5162	0.4306	0.8152	0.5035
VGG16 + UKSB+ GLOVE	300	0.7924	0.6814	0.5796	0.5044	0.4240	0.8058	0.4923
Inceptionv3 + UKSB +Dikkat	100	0.7972	0.6914	0.5915	0.5164	0.4313	0.8229	0.5140
VGG16 + UKSB + Dikkat	100	0.8054	0.7014	0.6030	0.5283	0.4359	0.8257	0.5203
Inceptionv3 + UKSB+ Dikkat+ GLOVE	100	0.8085	0.6980	0.5934	0.5138	0.4325	0.8205	0.5038
VGG16 + UKSB + Dikkat +GLOVE	100	0.8045	0.6926	0.5861	0.5051	0.4288	0.8164	0.4971
Inceptionv3 + UKSB +Dikkat	300	0.8042	0.6975	0.5967	0.5222	0.4342	0.8225	0.5140
VGG16 + UKSB + Dikkat	300	0.7974	0.6890	0.5871	0.5109	0.4292	0.8175	0.5057
Inceptionv3 + UKSB+ Dikkat+ GLOVE	300	0.8106	0.6977	0.5919	0.5122	0.4324	0.8201	0.5234
VGG16 + UKSB + Dikkat +GLOVE	300	0.7999	0.6848	0.5758	0.4962	0.4254	0.8121	0.4872








Çizelge 5. Her bir ürün özelliğinin altyazı oluşturma performansına etkisinin değerlendirmesi

Özellik Adı	B-1	B-2	B-3	B-4	M	R	C
Cinsiyet	0.5300	0.3791	0.2855	0.2246	0.2570	0.5362	0.2081
Ana Kategori	0.4911	0.3430	0.2519	0.1925	0.2391	0.5001	0.1838
Alt Kategori	0.5087	0.3619	0.2697	0.2087	0.2492	0.5227	0.2010
Eşya Türü	0.5239	0.3721	0.2752	0.2104	0.2560	0.5388	0.2082
Temel Renk	0.5803	0.4456	0.3511	0.2823	0.2952	0.5966	0.2698
Mevsim	0.5085	0.3607	0.2690	0.2080	0.2473	0.5155	0.1972
Yıl	0.5197	0.3747	0.2824	0.2209	0.2544	0.5306	0.2099
Kullanım Tipi	0.5005	0.3537	0.2610	0.1997	0.2445	0.5118	0.1901
Marka Adı	0.6862	0.5523	0.4369	0.3540	0.3538	0.7041	0.3605
Yaş Grubu	0.5191	0.3661	0.2716	0.2100	0.2503	0.5269	0.1945
Ücreti	0.5005	0.3526	0.2617	0.2015	0.2430	0.5043	0.1883

Deneyler sonucunda dikkat mekanizmasını kullanan sistemlerin diğerlerine göre daha iyi sonuçlar verdiği görülmüştür. Ayrıca model performansına etki derecesinin en yüksek özelliğın ise marka adı

olduğu gözlemlenmiştir. Ürün başlıklarında marka adlarının da yer alması nedeniyle bu sonucun normal olduğu düşünülmektedir. Marka adı özelliğinin ardından en etkin özelliklerin temel renk ve cinsiyet özellikleri olduğu da sonuçlarda görülmektedir. Çizelge 6' da tüm özelliklerin önerilen sisteme verilmesi sonucunda oluşturulan modelin performansı test kümesindeki bazı örnekler ile birlikte gösterilmiştir.

Çizelge 6. Eğitim sonrasında modelin test kümesinde doğru ve yanlış olarak tespit ettiği bazı örnekler (Yanlış tahmin edilen kelimeler kırmızı renkle ile gösterilmiştir)

Ürün Görüntüsü	Referans	Aday
	red tape men grey shoes	red tape men grey casual shoes
	chimp men teja blue tshirts	chimp men rajni blue tshirts
	grendha women kall brown flats	grendha women flori brown flats
	scullers for her check red shirt	scullers for her check red shirt
	stoln women black sunglasses	stoln women black sunglasses
	nike men solid black sweatshirts	nike men squad black sweatshirts
	arrow men grey check scarf	arrow men grey scarf
	murcia women white handbag	murcia women white handbag
	soliver mens black polo tshirt	soliver mens denim black tshirt

5. SONUÇLAR

Yapay zeka sistemleri insanların iş yükünü hafifleten farklı alandaki projelerde kullanılmaktadır. Bu çalışmanın amacı insan kaynaklı hataların zaman zaman karşılaşıldığı e-ticaret sistemlerine görüntü altyazısını otomatik oluşturan sistemlerinin adapte edilip edilmeyeceğinin ölçülmesi ve bu probleme eklenen farklı parametrelerin problemin başarısını hangi ölçülerde etkileyeceğinin saptanmasıdır.

Görüntü altyazısı oluşturma, bilgisayar görü ve doğal dil işleme tekniklerinin birlikte kullanılmasını gerektiren karmaşık bir sorundur. Modelinin performansı, veri kümesinin boyutu, açıklama oluşturmak için kullanılan kelime dağarcığı, görüntülerden öznitelik çıkarımı için kullanılan model, kullanılan ÖSA modeli ve modelde yer alan hiper parametreleri ayarlama gibi birçok faktöre bağlıdır.

Önerilen modelde çoğunluğu moda ürünlerinden oluşan FPID üzerinde deneyler çalışmalar gerçekleştirilmiştir. Deneyler sonucunda kodlayıcı kısmına eklenen özelliklerin sisteme dahil edilmesiyle 0.8106 BLEU-1 değeri elde edilmiştir. Deneysel değerlendirmelerde e-ticaret gibi farklı birçok parametrenin olduğu sistemlerde kodlayıcı kısmına verilen özelliklerin altyazısı oluşturma performansını oldukça yükselttiği görülmüştür. Çizelge 6’ da modelin test veri kümesinden rastgele seçilen veriler üzerindeki performansı gösterilmiştir.

Daha fazla verinin sistem performansını olumlu yönde etkileyeceği düşünülmektedir. Ayrıca FPID’ da bazı ürün görüntülerinin insan modeli üzerinde bazılarının ise modelsiz bir şekilde yer aldığı görülmüştür. Bölütleme teknikleri kullanılarak ürünlerin insan modelleri ile ayrıştırılmasının sonuçları olumlu yönde etkileyeceği düşünülmektedir. Gelecekte moda üzerine Türkçe görüntü altyazısı ve video altyazısı oluşturma çalışmalarının çalışmaların yapılması planlanmaktadır.

KAYNAKLAR

1. Yılmaz H., Sükman S. Bilgisayarlı Tomografi Görüntülerinden Kemik Dokunun Modellenmesi ve Fdm Yöntemiyle Baskısı, International Journal of 3D Printing Technologies and Digital Industry, Sayfa 227-235, 2019.
2. Turgut A., Temir A., Aksoy B., Özsoy K. Yapay Zekâ Yöntemleri ile Hava Sıcaklığı Tahmini için Sistem Tasarımı ve Uygulaması, International Journal of 3D Printing Technologies and Digital Industry, Sayfa 244-253, 2019.
3. Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., Luo, J., ve Bigham, J. P., “VizWiz Grand Challenge: Answering Visual Questions from Blind People”, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Pages 3608-3517, 2018.
4. Ling, H. ve Fidler, S., “Teaching Machines to Describe Images via Natural Language Feedback”, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Pages 5075-5085, 2017.
5. Karpathy, A., Joulin, A., ve Li, F.-F., “Deep Fragment Embeddings for Bidirectional Image Sentence Mapping”, NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2, Pages 1889–1897, 2014.
6. You, Q., Jin, H., Wang, Z., Fang, C., ve Luo, J., “Image Captioning with Semantic Attention, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pages: 4651-4659, 2016.
7. Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S., ve Bengio, Y., “Show, attend and tell: Neural image caption generation with visual attention”, 32nd International Conference on Machine Learning, ICML 2015, c. 3, Pages. 2048-2057, 2015.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., ve Polosukhin, I., “Attention is all you need”, Advances in Neural Information Processing Systems, Pages 5999-6009, 2017.
9. Farhadi, A., Hejrati, M., Sadeghi, M. A., Young, P., Rashtchian, C., Hockenmaier, J., ve Forsyth, D., “Every picture tells a story: Generating sentences from images”, European conference on computer vision, Pages 15-29, 2010.
10. Feng, Y. ve Lapata, M., “Automatic Caption Generation for News Images”, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 35, Pages 797-812, 2012.
11. Vinyals, O., Toshev, A., Bengio, S., ve Erhan, D., “Show and Tell: {A} Neural Image Caption Generator”, IEEE Conference on Computer Vision and Pattern Recognition, Pages 3156-3164, 2014.

12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., ve Wojna, Z., “Rethinking the Inception Architecture for Computer Vision”, *Computer Vision and Pattern Recognition*, 2015.
13. Sonmez, E. B., Yıldız, T., Yılmaz, B. D., ve Demir, A. E., “Türkçe dilinde görüntü altyazısı: veritabanı ve model”, *Gazi Üniversitesi Mühendis. Mimar. Fakültesi Dergisi*, Cilt 35, Sayı 4, Sayfa 2089-2100, 2019.
14. Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., ve Zhang, L., “Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering”, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Pages. 6077-6086, 2017.
15. Herdade, S., Kappeler, A., Boakye, K., ve Soares, J., “Image Captioning: Transforming Objects into Words”, *arXiv*, 2019.
16. Yang, X., Zhang, H., Jin, D., Liu, Y., Wu, C.-H., Tan, J., Xie, D., Wang, J., ve Wang, X., “Fashion Captioning: Towards Generating Accurate Descriptions with Semantic Rewards”, *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, c. 12358 LNCS, Pages. 1-17, 2020.
17. Li, X., Ye, Z., Zhang, Z., ve Zhao, M., “Clothes image caption generation with attribute detection and visual attention model”, *Pattern Recognit. Lett.*, Vol. 141, Pages 68-74, 2021.
18. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., ve Fei-Fei, L., “ImageNet: A Large-Scale Hierarchical Image Database” 2009 IEEE Conference on Computer Vision and Pattern Recognition, Pages 248-255, 2009.
19. He, K., Zhang, X., Ren, S., ve Sun, J., “Deep Residual Learning for Image Recognition”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Pages 770-778, 2015.
20. Simonyan, K. ve Zisserman, A., “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *CoRR, The 3rd International Conference on Learning Representations (ICLR2015)*, 2014.
21. Bahdanau, D., Cho, K., ve Bengio, Y., “Neural Machine Translation by Jointly Learning to Align and Translate”, *International Conference on Learning*, 2016.
22. Pennington, J., Socher, R., ve Manning, C. D., “Glove: Global Vectors For Word Representation”, *Proceedings Of The 2014 Conference on Empirical Methods In Natural Language Processing(EMNLP)*, Pages 1532-1543, 2014.
23. Papineni, K., Roukos, S., Ward, T., ve Zhu, W.-J., “BLEU: A Method For Automatic Evaluation Of Machine Translation”, *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Pages. 311-318, 2001.
24. Vedantam, R., Zitnick, C. L., ve Parikh, D., “CIDEr: Consensus-based Image Description Evaluation”, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Pages 4566-4575, 2015.
25. Banerjee, S. ve Lavie, A., “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, *Proceedings of the Second Workshop on Statistical Machine Translation*, Pages 228–231, 2005.
26. Lin, C.-Y., “ROUGE: A Package for Automatic Evaluation of Summaries”, *Proceedings of the ACL Workshop: Text Summarization Braches Out*, 2004