

Makale Özeti 2

Ilias Mantziaras

Mart 2025

1 Kullanılan Makale

A REVIEW OF TRANSFORMER-BASED MODELS FOR COMPUTER VISION TASKS: CAPTURING GLOBAL CONTEXT AND SPATIAL RELATIONSHIPS /2024 /arXiv

2 Özet

2.1 Transformerlar

Transformatörler, bilgisayarların resimleri ve kelimeleri anlamalarına yardımcı olan özel bilgisayar programlarıdır.Çok sayıda bilgiye bakmakta ve bunların nasıl bir araya geldiğini anlamakta çok iyidirler.Transformatörler ilk olarak bilgisayarların kelimeleri anlamalarına yardımcı olmak için kullanıldı, ancak artık bilgisayarların resimleri anlamalarına yardımcı olmak için de kullanılıyorlar.Bu, resimlerdeki nesneleri bulma, bir resimde ne olduğunu anlama ve hatta otonom arabaların yolu görmesine yardımcı olma gibi şeyler için yararlıdır.Transformatörler diğer bilgisayar programlarından farklıdır çünkü sadece küçük parçalarına bakmak yerine, bir kerede tüm resme bakabilirler.Ayrıca kendi kendilerine konuşabilir ve neyin önemli neyin önemli olmadığını anlayabilirler.Bu, onları karmaşık şeyleri anlamada çok iyi yapar.Birçok farklı transformatör türü vardır ve hepsi bilgisayarların resimleri ve kelimeleri daha iyi anlamalarına yardımcı olmak için kullanılır.

2.2 Bilgisayar Görüşü

Bilgisayar görüşü bilgisayarlar için bir süper güç gibidir. Resimlerde ve videolarda gördüklerini anlamalarına yardımcı olur. Resimlere bakıp nesneler, insanlar ve hayvanlar gibi şeyleri bulabilen modeller adı verilen özel araçlar vardır. Bu modeller bir resimde ne olduğunu bile anlayabilir. SMCA gibi bazı modeller bir resmin önemli kısımlarına odaklanabilir ve önemli olmayan şeyleri görmezden gelebilir. SWIN Transformer gibi diğerleri ise hem büyük hem de küçük şeyleri görmek için bir resme farklı mesafelerden bakabilir. Anchor DETR, bir

resimdeki nesneleri bulmaya yardımcı olmak için özel kutular kullanan bir modeldir. DEformable TRANSformer, bir resme bakıp şeylerin nasıl şekillendiğini ve nerede olduğunu anlayabilen başka bir modeldir. Tüm bu modeller özeldir çünkü bir resmin tamamına aynı anda bakabilir ve ne olduğunu anlayabilirler, bu da onların iyi kararlar almalarına yardımcı olur. Bunu yapmanın birçok yolu vardır ve bazıları transformatör adı verilen özel araçlar kullanır. Transformatörler, bilgisayarların bir resmin önemli kısımlarına odaklanmasına ve şeylerin birbirleriyle nasıl ilişkili olduğunu anlamasına yardımcı olur. Bir resmin küçük parçalarına ve büyük parçalarına bakabilir ve hatta şeylerin nasıl hareket ettiğini anlayabilirler. Bu, bilgisayarların resimlerdeki nesneleri bulma, bir videoda neler olduğunu anlama ve hatta araba sürme gibi şeyler yapmasına yardımcı olur. Birçok kişi transformatörleri daha iyi hale getirmek için çalışıyor, böylece bilgisayarların dünyayı daha iyi anlamasına yardımcı olabilirler. En iyi neyin işe yaradığını görmek için fikirlerini birçok farklı resim ve videoda test ediyorlar.

2.3 Obje Algılaması

Transformatör tabanlı modeller, resim veya videolardaki nesneleri bulma ve tanımlamanın bir yolu olan nesne algılama için kullanılır. Bu modeller, resmin tamamına bakma ve resmin farklı bölümlerinin nasıl ilişkili olduğunu anlama konusunda iyidir. Bunu yapmalarına yardımcı olmak için öz dikkat mekanizmaları gibi özel araçlar kullanırlar. Nesne algılama için kullanılan transformatör tabanlı modellere örnek olarak YOLOv5, Swin Transformer ve DETR verilebilir. Bu modeller, resim ve videolardaki nesneleri bulma konusunda iyidir ve hatta uzaktaki küçük nesneleri veya nesneleri bile bulabilirler. Ayrıca, bir resimde veya videoda ne olduğunu, örneğin içinde arabalar veya insanlar olup olmadığını anlama konusunda da iyidirler.