1.Özet

Yolov3’e dyaalı olarak, darknet içerisindeki Resblockların optimize edilip aynı genişlik ve yüksekliğe sahip ResNet birimi ile birleştirilmesi hedeflenmiştir. Ardından uzamsal bilgiyi iyileştirmek için early layer denen erken katmanda evrişim işlemi artırılarak tüm darknet yapısında iyileştirme yapılmıştır.

2.Veri Seti

UAV veriseti küçük objeler özel olarak insansız hava aracı perspektifinden toparlanmıştır, özel bir datasettir. Makalede UAV-viewed dataset olarak geçmektedir. İHA ile görüntülenen veri seti, normal veri setinde aynı seviyede performansı koruyarak küçük nesne algılamada belirgin performans artışı gösteriyor, bu da önerilen yöntemin farklı koşullara uyum sağladığı anlamına gelmektedir.

Bölüm 4.1 de datset tanıtılmış

İnternette UAV perspektif veri setinin bulunmaması nedeniyle, bu çalışma UAV123 ve kendi çekimleriyle toplanmış görüntüleri içerir. Eğitim verileri (3776 görüntü) ve test verilerine (630 görüntü) 6:1 oranında ayrılmış toplam 4406 görüntü içerir. Üç kural izlenerek eğitim/test verilerine ayrılırlar:

* Aynı videodaki görüntüler ayrık bir şekilde toplanır, bu da sabit bir aralıkla bir görüntü topladığımız anlamına gelir (her 5 sürekli görüntüden 1 görüntü topluyoruz);
* Aynı videodan alınan görüntüler ya eğitim verisi ya da test verisi olmalıdır;
* Aynı uçuştan birden fazla video varsa bunlar eğitim verisi ve test verisi olarak rastgele ayrılmalıdır. Veri kümesi, UAV-viewed olarak adlandırılır ve labelImg tarafından etiketlenir.

İnsan tespitine odaklanıldığı için, UAV-YOLO ilk önce VOC ve COCO'dan seçilen 8102 numune dahil 74.910 insan numunesi tarafından eğitilir. Ardından, İHA ile görüntülenen veri setine dayalı olarak transfer öğrenmesi uygulanır. Model, VOC/COCO'dan seçilen 5000 numune ve toplanan İHA görüntülemeli test veri seti tarafından test edilir.

3.Yöntem/Metot

İHA veriseti genel olarak arkaplan karmaşası içerebileceğinden ve hedef objelerin küçük olmasından dolayı, arkaplan karmaşası ve hedef ile kamera arasındaki mesafe dikkate alınarak üç kategoriye uzak, normal ve oyun olarak ayrılır ve eğitime bu şekilde dahil edilir. Veri artırma kullanılmıştır. Eğitim sırasındaanchor sayısı ve boyutu optimize olması için K-means gerçekleştirilir. Backbone network performansı normal konşullarda değişmeden tutarken küçük nesne algılamayı tahmin edecek şekilde optimize etmiştir.

Girdi boyutları 608x608 olarak boyutlandırılır, feature mapler 19x19x18 tensörler olarak darknet (yolov3 backbone) tarafından çıkarılır. Giriş görüntüleri 7x7ızgaralara (grid) bölünür, hedef objenin merkezi hangikaroya düşerse nesne tespitinden o karo sorumlu olmaktadır. Her döşeme için 2 bounding box oluşturulur ve bu kutular için confidence score tahmin edilir. Bu skorlar dedktörün kutunun o hedefi ne kadar içerdiğinden emin olmak için veri sunar. Pr(object)x IoU = confidence.

Herhangi bir obje yoksa güven skoru sıfır olmalıdır. Her sınırlayıcı kutu beş tahmin içerir kutucuğun sınırlarına göre kutunun merkezi (x,y), genişlik yükseklik (h,w), ve güven puanı. Her döşeme C koşullu sınıfı olarak Pr(Ci | Object) tahmin eder.

FR

Küçük cisimlerin hatalı tespit edilmesinin nedeni, sınırlı alıcı alandır. Bu sorunun etkisini azaltmak için YOLO'daki ResNet birimi ve omurga ağı geliştirilmiştir. YOLO'ya benzer şekilde UAV-YOLO, birden fazla sınırlayıcı kutuyu ve bu kutular için sınıf olasılıklarını aynı anda tahmin eden tek bir evrişimli ağ kullanır. Ayrıca tam görüntüler üzerinde eğitim alır ve algılama performansını doğrudan optimize eder

* UAV-YOLO'ya girdi olarak 608×608×3 olarak yeniden boyutlandırılan bir görüntü verildiğinde, girdinin öznitelikleri UAV-YOLO'nun omurga ağı tarafından çıkarılmaktadır.
* Hedeflerin farklı boyutlarını kesin olarak tespit etmek için, Şekil 1a'da y1, y2 ve y3 olarak ifade edilen 3 farklı ölçek ölçeği öngörülmüştür
* İHA görüntülemeli veri seti deneylerimizde, her ölçekte 3 kutu tahmin ettik; bu, 4 sınırlayıcı kutu koordinatı, 1 güven puanı ve 1 sınıf için tensörün N × N × [3 ˙ (4 + 1 + 1)] olduğu anlamına gelir. – insan tahmini için.
* Burada N, y1, y2 ve y3'ün özellik haritalarının boyutudur; sırasıyla 19, 38 ve 76'dır
* LeakRelu kullanılmıştır.

Özetle YOLOv3 ile İHA-YOLO arasındaki temel fark, omurga ağının yapısıdır. Önerdiğimiz RU\_2 nedeniyle, İHA-YOLO'daki omurga ağı daha derindir. Ek olarak, RU\_2, farklı RU\_1'lerde iki DBL'yi birleştirmek için ekstra kısayol bağlantısını artırdığından, YOLOv3'teki Res Biriminin tekrarı da UAV-YOLO'nunkinden farklıdır.

UAV-YOLO, YOLOv3'e dayalı olarak iki açıdan optimize edilmiştir:

* Aynı çıktı boyutuna sahip farklı Resblock'larda iki DBL'nin birleştirilmesiyle gerçekleştirilen Res Unit\_2'yi tasarlamak için kısayol bağlantısı gerçekleştirilir;
* Ağ yapısını derinleştirmek için her bir Res Biriminin tekrar süreleri değiştirilir. Bu iki optimizasyonu birleştirerek yalnızca alıcı dosyalamayı artırmakla kalmaz, aynı zamanda ağın semantik özellik çıkarma yeteneğini de geliştirir.

YOLOv3 ilk olarak sınıflandırılmadan toplanan İHA görüntülü veri seti ile eğitilmiştir. Dedektör, mAP sadece %51,41 ve IOU doğruluğunun %66,1 olduğu test veri setinde test edilmektedir. İHA görüntülü insan algılama performansını artırmak amacıyla, Şekil 3'te gösterildiği gibi YOLO eğitim yöntemi optimize edilmiştir.

diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 3

Çalışmada “uzak” örneklemler, boyutlarının küçük olması nedeniyle zor örneklemdir. Ağ, 5000 epoch ile "uzak" örnek tarafından tutulur.

4.Çıktı

UAV-YOLO üzerindeki deneysel sonuçlar ve diğer dedektörlerle karşılaştırmalar sunulmaktadır. i7-7820X 3.6 GHz CPU ve 64 GB RAM ile tek bir Nvidia GTX Titan XP GPU kullanarak karşılaştırma deneyleri yapılmaktadır

* Nesne algılama performansı, mAP ve zemin gerçeği ile öngörülen sınırlayıcı kutu (IOU) arasındaki örtüşme oranı ile değerlendirilir.
* Model yakınsamasını hızlandırmak için, ağırlık başlatma olarak ImageNet üzerindeki ağırlık eğitimi kullanılır.
* Sonuçlar Tablo 1'de gösterilmektedir. Orijinal İHA görüntülemeli veri seti eğitimi ile karşılaştırıldığında, sınıflandırılmış İHA görüntülemeli veri seti eğitiminin performansı önemli ölçüde artırabildiği görülebilir. Tablo 1'de gösterildiği gibi, mAP %51,41'den %90,88'e yükselmiştir.

masa içeren bir resim

Açıklama otomatik olarak oluşturuldu

* K-means ile anchor box optimize edilmiştir. 9 bağlantı kutusu ayarının IOU'da %80,59'a ulaşabilen en iyi performansı alabileceği bulunmuştur. Çapa9'un en iyi şekilde çalışmasının nedeni, alıcı alanı artırabilmesi ve hedef hakkında daha ayrıntılı bilgi alabilmesidir.
* "Uzak" örneklerle model yeniden eğitimi. Model, "normal" ve "oyunlar" koşullarında sınırlı azalan iyileştirme ile "uzak" koşuldaki performansı artırabilen "uzak" örneklerle yeniden eğitilmiştir. Tablo 1'de, mAP ve IOU'nun "uzak" örnekler için %1,05 ve %4,41 arttığı görülebilir; ancak, mAP "normal" ve "oyunlar" örnekleri için sırasıyla %0,01 ve %0,43 azalır
* "Uzak" örneklerdeki performansın çok iyileştiği ve diğer koşullardaki performansın neredeyse değişmeden kaldığı görülebilir. YOLOv3, SSD300 ve SSD512 dahil diğer dedektörlerle karşılaştırıldığında, UAV-YOLO'nun hem mAP'si hem de IOU'su çok artarak sırasıyla %66,25 ve %68,86'ya ulaşmıştır.

masa içeren bir resim

Açıklama otomatik olarak oluşturuldu

masa içeren bir resim

Açıklama otomatik olarak oluşturuldu

5.Sonuç

Bu makalede YOLOv3 tabanlı bir İHA perspektif nesne algılama yöntemi önerilmiştir. Küçük nesnelerin yanlış algılanmasına neden olan nedeni analiz edip bulunmuş ve yanlış algılama, sınırlı alıcı alanlarda iyileştirmeler önerilmiştir. Bu durumu azaltmak için, önce aynı genişlik ve yüksekliğe sahip iki ResNet birimini birleştirerek karanlık ağda Resblock'u optimize edilmektedir. Daha sonra, uzamsal bilgiyi zenginleştirmek için erken katmanda evrişim işlemi artırılarak tüm karanlık ağ yapısı iyileştirilir. Her ikisinin de alıcı alanı genişletebileceğini belirtmekte fayda var. Ayrıca, İHA perspektifi veya küçük nesne tespiti için özel olarak kullanılan İHA görüntülü veri setini de toplanmıştır. İHA ile görüntülenen veri setine dayalı olarak, eğitim verileri sınıflandırması, k-means anchor kutusu onayı ve sabit veri tutma dahil olmak üzere optimize edilmiş bir eğitim yöntemi de geliştiriyoruz. Diğer yöntemlerle karşılaştırıldığında temsili deneysel sonuçlar, yöntemimizin genel olarak çeşitli zorluklarda, özellikle küçük nesne algılamada iyi performans gösterdiğini göstermektedir.

