



YOLO-Based Deep Learning Approach for Turkish Sign Language Recognition

Ahmet İlisulu ^{1*}, Yakup Kutlu ¹

¹ İskenderun Teknik Üniversitesi, Bilgisayar Mühendisliği, İskenderun, Hatay, Türkiye.

ABSTRACT

We actually undertook this project based on a very simple need: to alleviate some of the difficulties hearing-impaired individuals experience when expressing themselves. While thinking about how to make TID more understandable, it made sense to incorporate artificial intelligence. YOLO models are already well known for their speed and practicality, so we went straight for them. We prepared a total dataset of 9,534 images. We added not only the basic signs we knew but also signs like “Hurry up” and “Help me,” which would be useful in emergencies in daily life. Because these are the things that are actually needed in real life. YOLOv11s was very fast but a bit weak in terms of accuracy. So it works, but not quite at the level we wanted. YOLOv8m was great in terms of accuracy, but it really took its toll on the system; it was slow and training took a long time. YOLOv11m, on the other hand, struck the perfect balance. It wasn't too slow or too cumbersome. It delivered very clean results with 0.993 mAP50 and 0.991 F1-Score. In short, we decided that YOLOv11m is the most logical choice for a real-time application. It's both fast and sufficiently accurate.

ARTICLE INFO

Received 17.11.2025

Accepted 21.12.2025

Publication Date 25.11.2025

Keywords:

Deep Learning, YOLO, Sign Language, Classification

Distributed Under CC-BY 4.0



Türk İşaret Dili Tanıma için YOLO Tabanlı Derin Öğrenme Yaklaşımı

ÖZET

Bu çalışmayı aslında çok basit bir ihtiyaçtan yola çıkarak yaptık: İşitme engelli bireylerin kendini ifade ederken yaşadığı sıkıntıları biraz olsun azaltmak. “TİD nasıl daha anlaşılır hâle gelebilir?” diye düşünürken işin içine yapay zekâyı katmak mantıklı geldi. YOLO modelleri de zaten hız ve pratiklik konusunda epey bilinen şeyler, o yüzden direkt onlara yöneldik. Toplamda 9.534 karelik bir veri seti hazırladık. Bu veri setine sadece bildiğimiz temel işaretleri değil, günlük hayatta acil durumlarda işe yarayacak “Acele et” ve “Yardım et” gibi işaretleri de ekledik. Çünkü gerçek hayatta asıl ihtiyaç duyulan şeyler bunlar oluyor. YOLOv11s çok hızlıydı ama doğruluk kısmında biraz zayıf kaldı. Yani iş görür ama tam istediğimiz seviyede değil. YOLOv8m doğruluk olarak harikaydı ama sistemin canını çıkarıyor resmen; ağır çalışıyor, eğitim uzun sürüyor. YOLOv11m ise tam dengede durdu. Ne çok yavaş, ne çok hantal. 0.993 mAP50 ve 0.991 F1-Score ile gayet temiz sonuç verdi. Kısacası gerçek zamanlı bir uygulama için en mantıklı seçeneğin YOLOv11m olduğuna karar verdik. Hem hızlı hem de yeterince doğru.

MAKALE BİLGİSİ

Received 17.11.2025

Accepted 21.12.2025

Publication Date 25.11.2025

Keywords:

Derin Öğrenme, YOLO, İşaret Dili, Sınıflandırma

Distributed Under CC-BY 4.0



GİRİŞ

İnsanlık tarihi boyunca iletişim farklı yöntemlerle sağlanmıştır; mağara duvarlarına çizilen resimler, Afrika yerlilerinin çıkardıkları sesler ve Kızılderililerin duman yoluyla iletişimleri buna örnektir (Yüksel, 1994). İnsanlar arasında en yaygın kullanılan iletişim türü, duyma ve konuşma esasına dayanan sözlü iletişimdir. Ancak konuşma ya da işitme ile ilgili bir engeli olan bireyler, sözlü iletişimin hakim olduğu dünyada ciddi bariyerlerle karşılaşmaktadır. Dünya Sağlık Örgütü'nün (WHO) güncel raporlarına göre, küresel ölçekte 430 milyondan fazla insan rehabilitasyon gerektiren düzeyde işitme kaybına sahiptir ve bu sayının nüfusun yaşlanmasıyla birlikte 2050 yılında 700 milyonu aşacağı öngörülmektedir. (World Health Organization [WHO], 2024). Bu veriler, işitme engeline yönelik çözümlerin sadece bireysel değil, küresel bir zorunluluk olduğunu göstermektedir. Özellikle eğitimleri oldukça zor olan özel gereksinime sahip bireylerde yöntemlerde farklı olmaktadır (Mermertaş ve Ark. , 2022).

İşitme kayıplı bireylerin bir kısmı işitme cihazı veya koklear implant ile işitsel-sözel iletişim kurabilmektedir. Bununla birlikte geç tanılanan, cihazlardan fayda göremeyen ya da kültürel bir tercih olarak işaret dilini benimseyen bireyler, iletişim aracı olarak işaret dilini kullanmaktadır. Akademik çalışmalarda, işaret dilini öncelikli iletişim dili olarak benimseyen ve kendilerini Sağır toplumuna ait hissedenen bireyler için 'Sağır' (büyük S ile), tıbbi olarak işitme kaybı olan bireyler için ise 'sağır' (küçük s ile) kavramı kullanılmaktadır. Türk İşaret Dili (TİD), sadece el hareketlerinden ibaret olmayıp; kaş, göz, baş ve gövde hareketleri gibi "manuel olmayan işaretleyiciler" (non-manual markers) içeren, kendine özgü grameri ve sözdizimi olan doğal bir dildir (Piştav Akmeşe, 2016).

Türkiye İstatistik Kurumu'nun 2010 yılı araştırmasına göre, sağır bireylerin %31,6'sı okur-yazar değildir (TÜİK, 2010). Bu durum, yazılı ve sözlü iletişimde ciddi engeller yarattığından işaret dilinin önemini artırmaktadır. İşaret dili ilk defa 5378 Sayılı kanunla ülkemizde yasal statü kazanmıştır. Kanunun hemen akabinde gerekli mevzuat çalışmaları yapılarak işitme engelli bireylerin iletişim engellerine çözüm sunmak üzere işaret dili tercümanları istihdam edilmiştir (TBMM,2024).

MATERYAL ve YÖNTEM

Bilgisayarlı görü (computer vision) alanında temel bir görev olan nesne algılama, son yıllarda geliştirilen algoritmalarla hız ve doğruluk açısından büyük ilerlemeler kaydetmiştir. Bu alandaki en önemli kırılma noktalarından biri, "You Only Look Once" (YOLO) algoritmasının ortaya çıkışıdır. İlk kez 2015 yılında Redmon ve arkadaşları tarafından tanıtılan YOLO, nesne algılamayı çok aşamalı bir süreçten tek aşamalı (one-stage) bir regresyon problemine dönüştürerek devrim yaratmıştır (Redmon ve ark., 2016).

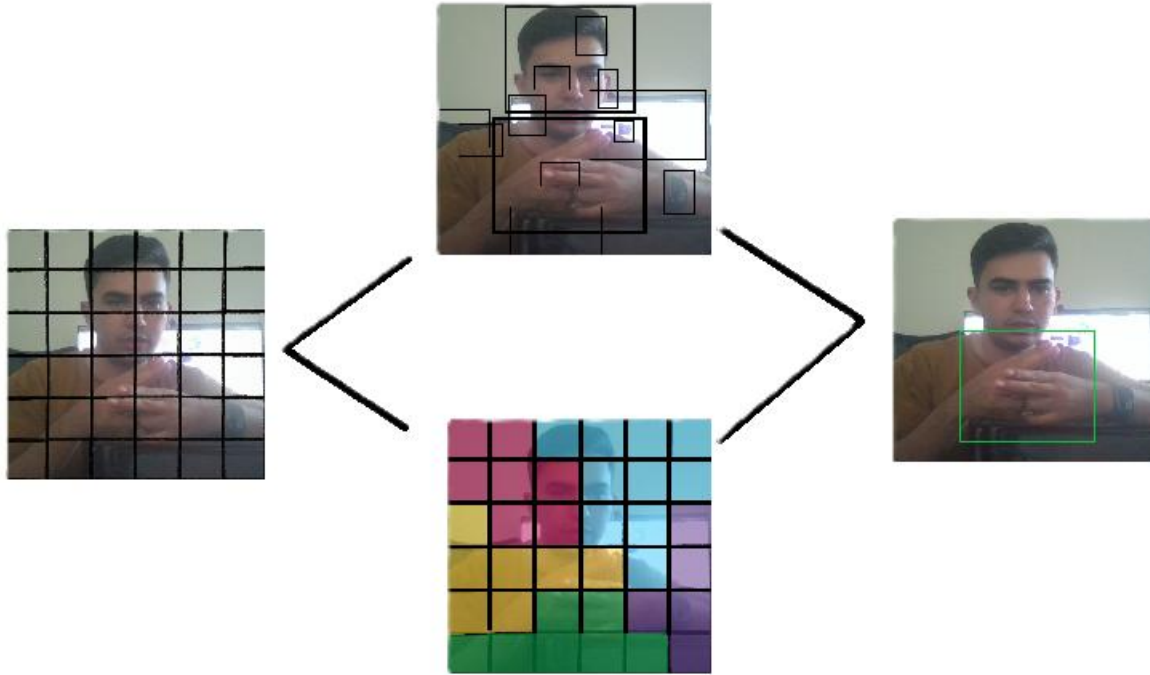
YOLO Serisinin Tarihsel ve Mimari Gelişimi

YOLO'nun evrimi, hız ve doğruluk arasındaki dengeyi optimize etmeye yönelik sürekli bir arayışı temsil eder.

Erken Dönem (v1 - v3): YOLOv1, saniyede 45 kare (fps) işleme hızıyla gerçek zamanlı algılamanın kapısını aralamış olsa da, özellikle küçük nesnelerin tespitinde ve yerleştirmede zorluklar yaşamıştır. Bu sorunu aşmak için geliştirilen YOLOv2 (YOLO9000), Darknet-19 omurgasını kullanarak çözünürlüğü artırmış ve 9000'den fazla nesne kategorisini tanıyabilir hale gelmiştir (Li ve Yang, 2018). Ardından gelen YOLOv3, daha derin bir yapı olan Darknet-53 omurgası ve artık (residual) bağlantılar ile sahneye çıkmıştır. YOLOv3'ün en önemli yeniliği, farklı boyutlardaki nesneleri (özellikle küçük olanları) daha iyi algılamak için çok ölçekli tahmin yapabilen Özellik Piramit Ağları (FPN) benzeri bir yapı kullanmasıdır (Kim ve ark., 2018).

Orta Dönem ve Optimizasyonlar (v4 - v10): Serinin devamında, hesaplama verimliliği ve kullanıcı dostu yapılar ön plana çıkmıştır. YOLOv4 ve YOLOv5, CSP (Cross Stage Partial) ağırları ve mozaik veri artırma teknikleri ile hem hızı hem de doğruluğu artırmıştır (Ultralytics, 2020). YOLOv6'dan itibaren endüstriyel uygulamalara odaklanılmış, YOLOv10 ile birlikte Maksimum Olmayan Baskılama (NMS) işlemine duyulan ihtiyaç ortadan kaldırılarak gecikme süreleri (latency) minimize edilmiştir (Wang ve ark., 2024).

Modern Dönem ve Dikkat Mekanizmaları (v11 - v12): En güncel sürümler olan YOLOv11 ve YOLOv12, mimaride köklü değişikliklere gitmiştir. Özellikle YOLOv12, dikkat merkezli (attention-centric) bir tasarım benimseyerek Alan Dikkat (A2) modüllerini entegre etmiştir. Bu sayede, önceki CNN tabanlı modellerin sınırlamaları aşarak nesne kontur tanımında üstün performans elde edilmiştir (Tian ve ark., 2025).



Resim 1. Yolo Mimarisi.

BULGULAR

Bu çalışmada yürütülen deneysel süreç, açık kaynaklı Türk İşaret Dili (TİD) veri setinin kullanılması ve bu veri setine bazı özel işaretlerin eklenmesiyle yapılandırılmıştır. Eğitim ve test aşamalarında veriler, yaygın literatür uygulamalarına uygun olarak %80 eğitim ve %20 test olacak şekilde ayrılmıştır. Modellemenin gerçekleştirilmesinde YOLOv11s, YOLOv11m ve YOLOv8m

mimarileri tercih edilmiş; tüm modeller aynı donanım konfigürasyonu (NVIDIA GPU altyapısı) üzerinde eğitilerek karşılaştırmalı bir değerlendirme yapılmıştır.

Elde edilen deneysel bulgular, YOLOv11s modelinin temel düzeyde başarılı bir performans sunduğunu ve özellikle hız açısından avantajlı olduğunu göstermektedir. YOLOv11m modeli, doğruluk ve güvenilirlik bakımından en dengeli sonuçları üretmiş ve bütüncül performans ölçütlerinde diğer modellerin üzerinde bir başarı göstermiştir. Buna karşılık YOLOv8m modeli yüksek doğruluk değerlerine ulaşmasına rağmen, eğitim süresi ve hesaplama maliyeti açısından daha yüksek kaynak gerektirmiştir. Bu sonuçlar ışığında, Türk İşaret Dili tanıma görevinde optimum doğruluk-hız dengesi göz önünde bulundurulduğunda, YOLOv11m mimarisi en uygun seçenek olarak öne çıkmaktadır.

Çalışmada kullanılan veri seti toplam 9.534 görüntüden oluşmaktadır. Veri setinde Anne, Arkadaş, Baba, Dur, Ev, Evet, Hayır, Kardeş, Merhaba, Nasıl, Nerede, Özür Dilemek, Tamam, Telefon, Teşekkürler, Tuvalet, Yemek, İçmek, İyi ve Kötü olmak üzere 20 temel Türk İşaret Dili işareti yer almaktadır. Ayrıca veri setinde bulunmayan “Acele et” ve “Yardım et” işaretleri çalışmanın gereksinimleri doğrultusunda tarafımızca örneklendirilerek veri setine dâhil edilmiştir.

Nesne tespiti için derin öğrenme tabanlı YOLO algoritması tercih edilmiştir. YOLO, nesne tespitini tek aşamada gerçekleştiren ve gerçek zamanlı işlem kapasitesi sunan bir mimari olduğu için işaret dili tanıma uygulamalarında etkin bir çözüm sağlamaktadır. Bu kapsamda üç farklı model eğitilmiş ve birbirleriyle karşılaştırılmıştır (Roboflow 2025), (Redmon ve ark., 2016)).:

- YOLOv11s
- YOLOv11m
- YOLOv8m

Model performanslarının değerlendirilmesinde mAP (mean Average Precision), Precision, Recall ve F1-Score gibi yaygın performans ölçütleri kullanılmıştır.



Resim 2. Veri setine tarafımızca eklenen işaretlerden örnek.

SONUÇ VE TARTIŞMA

Elde edilen sonuçlar genel olarak değerlendirildiğinde, YOLOv11m modelinin Türk İşaret Dili işaretlerinin otomatik tanınması görevinde en yüksek doğruluk ve istikrarı sağladığı görülmektedir. Modelin yüksek mAP, Precision, Recall ve F1-Score değerleri, özellikle gerçek zamanlı işaret dili

tanıma uygulamalarında güçlü bir aday olduğunu ortaya koymaktadır. Bu bağlamda, hem performans hem de kaynak verimliliği açısından YOLOv11m modelinin TİD tanıma sistemlerinde etkin bir çözüm sunabileceği sonucuna ulaşılmıştır.

DeneySEL Sonuçlar

Tablo 1. YOLO modelleri için test sonuçları.

Model	mAP50	mAP50-95	Precision (macro)	Recall (macro)	F1-Score (macro)
YOLOv11s	0.844	0.617	0.852	0.778	0.813
YOLOv11m	0.993	0.848	0.985	0.997	0.991
YOLOv8m	0.993	0.839	0.976	0.990	0.983

Gelecekte, böyle bir sistem tüm Türk İşaret Dili sembollerini kapsayacak şekilde genişletilebilir. Ayrıca, artırılmış gerçeklik gözlükleri ve mobil cihazlara entegre edilerek işitme engelli bireylerin günlük hayatta kesintisiz iletişim kurmaları mümkün olabilir. En büyük zorluk, bu tür sistemlerin donanım maliyetleridir. Ancak teknolojik gelişmeler ile bu maliyetlerin azalması, sistemin yaygınlaşmasına katkı sağlayacaktır.

KAYNAKÇA

- Li, R., & Yang, J. (2018). Improved YOLOv2 object detection model. In 2018 6th International Conference on Multimedia Computing and Systems (ICMCS). IEEE. <https://doi.org/10.1109/ICMCS.2018.8525895>
- Kim, K.-J., Kim, P.-K., Chung, Y.-S., & Choi, D.-H. (2018). Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE. <https://doi.org/10.1109/AVSS.2018.8639438>
- Mermertaş, F., Kapudere, M., Yerdoğan, M., Yenigün, B. (2022). BA-SA-RA Method-Based Educational Application for Special Students. *Journal of Artificial Intelligence with Applications*, 3(1), 20-23
- Piştav Akmeşe, P. (2016). Türk işaret dili (TİD) eğitimi ve yüksek öğretim programlarında işaret dili dersi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 17(3), 341–361.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html
- Roboflow. (2025). Türk İşaret Dili. Roboflow Universe. <https://universe.roboflow.com/proje-qtjgs/turk-isaret-dili>
- Ultralytics. (n.d.). Ultralytics documentation. Retrieved from <https://docs.ultralytics.com/>

- Türkiye İstatistik Kurumu. (2010), <https://tuik.gov.tr>
- Türkiye Büyük Millet Meclisi. (2024).(Dönem 28, Yasama Yılı2).
<https://cdn.tbmm.gov.tr/KKBSPublicFile/D28/Y2/T7/WebOnergeMetni/ad7e3eb3-09be-4fc7-ba3a-affee920ca20.pdf>
- Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. arXiv. <https://arxiv.org/abs/2502.12524>
- World Health Organization. (2024, 1 Mart). *Deafness and hearing loss*. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. In *Computer Vision – ECCV 2024* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-72751-1_1
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection. In *Advances in Neural Information Processing Systems, 37* (NeurIPS 2024). https://proceedings.neurips.cc/paper_files/paper/2024/hash/c34ddd05eb089991f06f3c5dc36836e0-Abstract-Conference.html
- Yüksel, H. (1994). *Bireylerarası İletişime Giriş*. Eskişehir: Anadolu Üniversitesi Yayınları
- Zhang, Y., Sun, Y., Wang, Z., & Jiang, Y. (2023). YOLOv7-RAR for urban vehicle detection. *Sensors*, 23(4), 1801. <https://doi.org/10.3390/s23041801>