KOMPYUTER KO‘RISH SOHASIDA OBYEKTLARNI ANIQLASH: YOLO YONDASHUVI VA SAMARADORLIK

[1]To‘xtasinov Mumtozali To‘lqinaliyevich, [2]Musojonov Xumoyun Muhammadayubxon o‘g‘li
[1](dotsent, texnika fanlari nomzodi; Namangan Davlat Texnika Universiteti; “Raqamli texnologiyalar” kafedrasi katta ilmiy xodimi
[2](doktorant; Namangan Davlat Texnika Universiteti; “Raqamli texnologiyalar” kafedrasi doktoranti
https://doi.org/10.5281/zenodo.17911902

Annotatsiya. *Ushbu maqola kompyuter ko‘rish (computer vision) sohasidagi* ***obyektlarni aniqlash (object detection)*** *vazifasini, uning algoritmlari va eng mashhur modellaridan biri —* ***YOLO (You Only Look Once)*** *ni tahlil qiladi. Maqolada single-shot(bir martalik) va two-stage(ikki martalik) detektorlar, baholash metrikalari (IoU, AP, mAP), YOLO arxitekturasi, uning ishlash prinsipi va samaradorligini oshiruvchi usullar (NMS-) muhokama qilinadi. Tadqiqot real vaqt rejimida obyektlarni aniqlashga yo‘naltirilgan ilovalar uchun foydali bo‘ladi.*

Kalit so‘zlar: *obyektlarni aniqlash, yolo, computer vision, single-shot(bir martalik) detektor, two-stage(ikki martalik) detektor, bounding box, average precision, non-maximum suppression.*

KIRISH: Obyektlarni aniqlash (Object detection) — kompyuter ko‘rish (computer vision) sohasida muhim va keng qo‘llaniladigan vazifalardan biridir [1]. Ushbu vazifa tasvir ichida qiziqish obyekti (region of interest) joylashgan hududni aniqlash va uni oddiy tasvir tasniflagichlari kabi turkumlash (klassifikatsiya qilish) bilan shug‘ullanadi. Bitta tasvirda bir nechta turli obyektlar bo‘lishi mumkin, shu sababli obyektlarni aniqlash oddiy tasvirni tasniflashga qaraganda ancha murakkab masala hisoblanadi [2].

YOLO (You Only Look Once) modeli obyektlarni aniqlashda tezligi va aniqligi bilan mashhur bo‘lib, birinchi marta Joseph Redmon va boshqalar tomonidan 2016-yilda taqdim etilgan. YOLO bir necha bosqichlardan o‘tgan bo‘lib, uning eng so‘nggi versiyalaridan biri — YOLOv7. Ushbu model boshqa obyekt aniqlash algoritmlaridan farq qiladigan yondashuvni qo‘llaydi va real vaqt rejimida ishlash imkonini beradi.

Obyektlarni aniqlash – bu tasvirlar yoki videolarda joylashgan obyektlarni aniqlash va ularning joylashuvini belgilash bilan shug‘ullanuvchi kompyuter ko‘rish vazifasidir. Bu vazifa kuzatuv tizimlari, o‘z-o‘zidan boshqariladigan (avtonom) avtomobillar, robototexnika kabi ko‘plab ilovalarda muhim rol o‘ynaydi.

Obyektlarni aniqlash algoritmlari umumiy tarzda, kiruvchi tasvirning neyron tarmoqdan necha marta o‘tishiga qarab ikki asosiy turga bo‘linadi: bir martalik (single-shot) detektorlar va ikki bosqichli (two-stage) detektorlar. Bir martalik detektorlar tasvirni faqat bir marta tarmoqdan o‘tkazadi va obyektlarning mavjudligi hamda joylashuvini bashorat qiladi. Bu yondashuv hisoblash jihatdan samarali bo‘lib, real vaqt ilovalari uchun qulay hisoblanadi, ammo kichik obyektlarda aniqligi pastroq bo‘lishi mumkin.

Ikki bosqichli detektorlar esa tasvirni ikki marta tarmoqdan o‘tkazadi. Birinchi bosqichda tasvirdagi obyektlar bo‘lishi mumkin bo‘lgan hududlar (region proposals) aniqlanadi, ikkinchi

bosqichda esa ushbu hududlar aniqlashtiriladi va yakuniy bashoratlar chiqariladi. Bu yondashuv aniqlikni oshiradi, ammo hisoblash xarajati yuqoriroq bo‘ladi.

Obyektlarni aniqlash muammosini chuqur o‘rganish (deep learning) yordamida hal qilish bo‘yicha ilk muvaffaqiyatli urinishlardan biri 2014-yilda Microsoft Research’da Ross Girshick va uning jamoasi tomonidan ishlab chiqilgan R-CNN (Regions with CNN features) modeli bo‘lgan [2]. Ushbu model tasvirlarda obyektlarni aniqlash va ularning joylashuvini belgilash uchun soha taqdimi **(**region proposal**)** algoritmlari va konvolyutsion neyron tarmoqlar **(**CNN**)** kombinatsiyasidan foydalangan.

YOLO modeli esa butunlay boshqa yondashuvni taklif qiladi. U yakuniy (end-to-end) neyron tarmog‘idan foydalanib, bir vaqtning o‘zida cheklov soxalari (bounding boxes) va sinf ehtimolliklarini (class probabilities) bashorat qiladi. Shu bilan u real vaqt rejimida yuqori aniqlik bilan obyektlarni aniqlash imkonini beradi.



1-rasm. Bir bosqichli va ikki bosqichli detektorlar

Bir martalik va ikki bosqichli obyekt aniqlash**:** bir martalik (single-shot) obyekt aniqlash usuli kiruvchi tasvirni faqat bir marta neyron tarmoqdan o‘tkazadi va undagi obyektlarning mavjudligi hamda joylashuvini bashorat qiladi. Ya’ni, butun tasvir bitta o‘tishda qayta ishlanadi, bu esa hisoblash jihatdan samaradorlikni oshiradi. Shu bilan birga, single-shot detektorlar kichik obyektlarni aniqlashda unchalik samarali bo‘lmasligi mumkin va boshqa usullarga nisbatan aniqligi biroz pastroq bo‘ladi. Shunga qaramay, bu algoritmlar resurslari cheklangan qurilmalarda real vaqt rejimida ishlash uchun juda qulaydir.

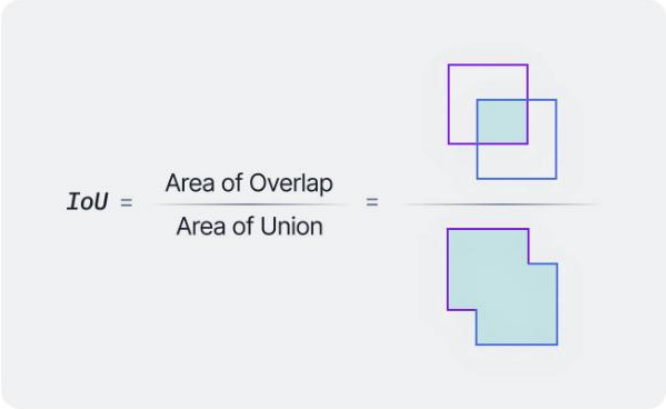
YOLO modeli — tasvirni qayta ishlash uchun to‘liq konvolyutsion neyron tarmog‘idan (FCNN) foydalanuvchi bir martalik detektordir. Keyingi bo‘limda biz YOLO modelini yanada chuqurroq ko‘rib chiqamiz.

Ikki bosqichli (two-stage) obyekt aniqlash usuli esa kiruvchi tasvirni ikki marta tarmoqdan o‘tkazadi, bunda obyektlarning mavjudligi va joylashuvi haqidagi bashoratlar bosqichma-bosqich amalga oshiriladi. Birinchi bosqichda tasvirdagi obyektlar bo‘lishi mumkin bo‘lgan ehtimoliy hududlar (proposals) aniqlanadi. Ikkinchi bosqichda esa ushbu takliflar aniqlashtiriladi va yakuniy bashoratlar chiqariladi.

Ikki bosqichli detektorlar bir martalik detektorlarga nisbatan aniqligi yuqori bo‘lsa-da, hisoblash xarajati ham ko‘proq bo‘ladi. Shu bois single-shot va two-stage yondashuvlaridan qaysi birini tanlash ilovaning talab va cheklovlariga bog‘liq bo‘ladi. Umuman olganda, real vaqt talab qilinadigan dasturlar uchun bir martalik detektorlar qulay, aniqlik yuqori bo‘lishi kerak bo‘lgan ilovalar uchun esa ikki bosqichli detektorlar mos keladi.

Obyekt aniqlash modellari samaradorligini baholash uchun standart miqdoriy metrikalardan foydalaniladi. Eng keng tarqalganlari — Intersection over Union **(**IoU**)** va Average Precision **(**AP**)** [3].

Intersection over Union **(**IoU**)** obyekt aniqlash modellarida joylashuv aniqligini o‘lchash va xatolarni hisoblash uchun ishlatiladi. Bashorat qilingan va haqiqiy bounding boxlar (chegara soxalari) o‘rtasidagi IoU’ni hisoblash uchun, avval bir xil obyektga tegishli ikkita bounding box kesishgan hududi (Intersection) aniqlanadi [3]. Shundan so‘ng, ikkita bounding box qoplagan umumiy maydon (Union) hisoblanadi. Intersection maydonining Union maydoniga nisbati obyekt qanchalik aniq bashorat qilinganini ko‘rsatadi va bashorat qilingan bounding box asl bounding boxga qanchalik yaqinligini baholashda muhim ko‘rsatkich hisoblanadi.



2-rasm. O‘rtacha aniqlik

O‘rtacha Aniqlik (Average Precision, AP) — model tomonidan qilingan bashoratlar to‘plami uchun aniqlik (precision) va chaqiruv (recall) o‘rtasidagi bog‘liqlik grafigi ostidagi maydon (AUC) orqali hisoblanadi. Recall (chaqiruv) — model ma’lum bir klass bo‘yicha nechta mavjud haqiqiy belgi (label)dan nechta topa olganini bildiradi, ya’ni klass bo‘yicha mavjud obyektlar soniga nisbatan to‘g‘ri aniqlangan obyektlar ulushi. Precision (aniqlik) — modelning o‘sha klass bo‘yicha qilgan bashoratlari ichida nechta to‘g‘ri aniqlangan (true positive) obyektlar ulushi. Recall va precision o‘rtasida muayyan almashuv (trade-off) mavjud bo‘ladi. Tasniflash threshold (chegarasi) o‘zgartirilganda ushbu ikki qiymat grafigi chiziladi va hosil bo‘ladigan precision-recall curve ostidagi maydon — o‘sha klass uchun Average Precision qiymatini beradi. Modeldagi barcha klasslar bo‘yicha olingan AP qiymatlarining o‘rtachasi mean Average Precision (mAP) deb ataladi.

Obyekt aniqlashda (object detection) esa precision va recall faqat klassni to‘g‘ri topish uchun emas, balki bounding box (chegara qutisi) bashoratining to‘g‘riligini baholash uchun ishlatiladi. Bu yerda qaror qabul qilish mezoni sifatida IoU (Intersection over Union) qo‘llanadi:

IoU > 0.5 bo‘lsa → bu ijobiy (positive) bashorat hisoblanadi.

IoU < 0.5 bo‘lsa → bu salbiy (negative) bashorat hisoblanadi.

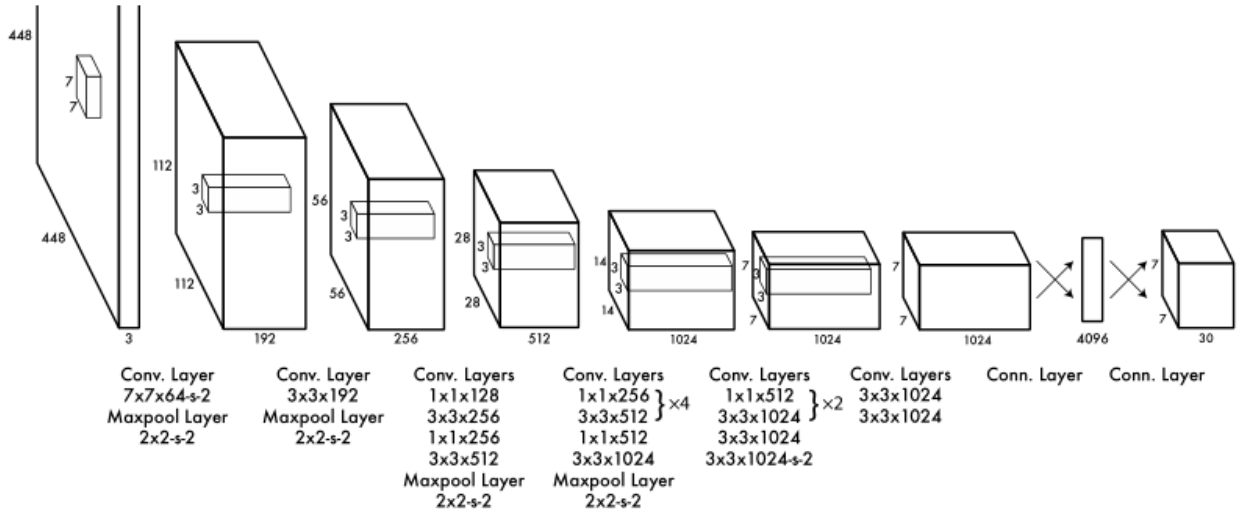


3-rasm. YOLO ni rivojlanish bosqichlari

You Only Look Once (YOLO) yakuniy (end-to-end) neyron tarmog‘idan foydalanishni taklif qiladi, u bir vaqtning o‘zida cheklov qutilari (bounding boxes) va sinf ehtimolliklarini (class probabilities) bashorat qiladi. Bu avvalgi obyekt aniqlash algoritmlaridan farq qiladi, ular aniqlash vazifasini bajarish uchun klassifikatorlardan qayta foydalanishgan. Obyekt aniqlashga butunlay boshqa yondashuvni qo‘llab, YOLO eng so‘nggi natijalarga erishdi va real vaqt rejimidagi boshqa obyekt aniqlash algoritmlarini ancha ortda qoldirdi. Masalan, Faster RCNN kabi algoritmlar Region Proposal Network yordamida qiziqish hududlarini aniqlab, keyin ushbu hududlarda alohida tanib olish (recognition) jarayonini bajaradi, YOLO esa barcha bashoratlarini bitta to‘liq bog‘langan (fully connected) qatlam yordamida bajaradi. Region Proposal Network’ni ishlatadigan usullar bir xil tasvir uchun bir nechta iteratsiyalarni bajaradi, YOLO esa bitta iteratsiya bilan kifoya qiladi. 2015-yilda YOLO dastlab chiqarilganidan beri bir nechta yangi versiyalari taklif qilindi, ularning har biri avvalgisini yaxshilash va rivojlantirishga asoslangan [4].

YOLO arxitekturasi. YOLO algoritmi tasvirni (image) kirish sifatida oladi va keyin tasvirdagi obyektlarni aniqlash uchun oddiy chuqur konvolyutsion neyron tarmog‘idan (CNN) foydalanadi.

YOLO ning asosini tashkil qiluvchi CNN modelining arxitekturasi quyida ko‘rsatilgan.



4-rasm. Arxitektura

Bizning aniqlash tarmog‘imizda 24 ta konvolyutsion qatlam mavjud bo‘lib, ularni 2 ta to‘liq bog‘langan (fully connected) qatlamlar kuzatadi. Navbatma-navbat 1 × 1 konvolyutsion qatlamlar

oldingi qatlamlardan olingan xususiyatlar fazosini (feature space) kamaytiradi. Biz konvolyutsion qatlamlarni ImageNet klassifikatsiya vazifasida yarim o‘lchamda (224 × 224 kirish tasviri) oldindan o‘qitamiz (pretrain) va keyin aniqlash uchun o‘lchamni ikki barobar oshiramiz.

Modelning birinchi 20 konvolyutsion qatlami ImageNet yordamida oldindan o‘qitiladi (pre-trained) va bu jarayonda vaqtincha o‘rtacha pooling (average pooling) va to‘liq bog‘langan (fully connected) qatlam qo‘shiladi. Keyin ushbu oldindan o‘qitilgan model aniqlash vazifasini bajarish uchun o‘zgartiriladi, chunki avvalgi tadqiqotlar shuni ko‘rsatdiki, oldindan o‘qitilgan tarmoqqa qo‘shimcha konvolyutsion va bog‘langan qatlamlarni qo‘shish natijani yaxshilaydi. YOLO ning yakuniy to‘liq bog‘langan qatlami sinf ehtimolliklarini va cheklov qutisi (bounding box) koordinatalarini bashorat qiladi.

YOLO kirish tasvirini S × S tarmoq (grid)ga bo‘ladi. Agar obyekt markazi tarmoq hujayrasiga tushsa, o‘sha hujayra shu obyektni aniqlash uchun mas’ul bo‘ladi. Har bir tarmoq hujayrasi B ta cheklov qutisi va ularning ishonch (confidence) ballarini bashorat qiladi. Ushbu ishonch ballari model qutida obyekt borligiga qanchalik ishonchi borligini va bashorat soxasi qanchalik aniq ekanligini aks ettiradi.

YOLO har bir tarmoq obyekti uchun bir nechta cheklov qutilarini bashorat qiladi. Mashg‘ulot (training) vaqtida biz faqat bitta cheklov qutisi bashorat qiluvchisini har bir obyekt uchun mas’ul qilishni xohlaymiz. YOLO obyektni bashorat qilishda “mas’ul” bo‘ladigan bashorat qiluvchini tanlaydi va bu tanlov joriy eng yuqori IOU (Intersection over Union)ga ega bo‘lgan bashoratga asoslanadi. Bu cheklov sozasi bashorat qiluvchilar o‘rtasida ixtisoslashuvni (specialization) keltirib chiqaradi. Har bir bashorat qiluvchi ma’lum o‘lchamlar, nisbatlar yoki obyekt sinflarini bashorat qilishda yaxshilanadi va natijada umumiy recall ko‘rsatkichi oshadi.

YOLO modellarida ishlatiladigan muhim usullardan biri — non-maximum suppression (NMS). NMS post-prosessing qadamidir va obyekt aniqlashning aniqligi va samaradorligini oshirish uchun ishlatiladi [4]. Obyekt aniqlashda bir tasvirdagi bitta obyekt uchun bir nechta cheklov qutilari yaratilishi odatiy hol. Ushbu qutilar bir-birini qoplash yoki turli joylarda joylashishi mumkin, lekin barchasi bir xil obyektni ifodalaydi. NMS ortiqcha yoki noto‘g‘ri cheklov qutilarini aniqlash va olib tashlash, hamda har bir obyekt uchun faqat bitta cheklov qutisini chiqarish uchun ishlatiladi.

XULOSA

Obyektlarni aniqlash (Object Detection) kompyuter ko‘rish sohasida muhim vazifa bo‘lib, tasvir va videolardagi obyektlarning mavjudligi va joylashuvini aniqlashga xizmat qiladi. Bir martalik (single-shot) va ikki bosqichli (two-stage) detektorlar bu vazifani bajarishda turli yondashuvlarni taklif qiladi: bir martalik detektorlar hisoblash jihatdan samarali bo‘lib, real vaqt ilovalari uchun qulay, ikki bosqichli detektorlar esa aniqligi yuqori bo‘lib, hisoblash xarajati ko‘proq.

Obyekt aniqlash modellarining samaradorligini baholashda Intersection over Union (IoU) va Average Precision (AP) kabi metrikalardan foydalaniladi. IoU *bounding box*larning aniqligini, AP esa precision va recall o‘rtasidagi bog‘liqlikni baholash orqali modelning umumiy ishlashini o‘lchaydi.

YOLO (You Only Look Once) modeli esa an’anaviy algoritmlardan farqli ravishda, yakuniy neyron tarmog‘i yordamida bir vaqtning o‘zida cheklov qutilari va sinf ehtimolliklarini bashorat qilib, real vaqt rejimida yuqori aniqlik bilan ishlash imkonini beradi. Shu sababli YOLO modeli

hozirgi kunda tezkor va samarali obyekt aniqlash sohasida eng ommabop yondashuvlardan biri sifatida e'tirof etiladi.

ADABIYOTLAR

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *CVPR*.
2. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR*.
3. Lin, T.-Y., Maire, M., Belongie, S., et al. (2014). Microsoft COCO: Common Objects in Context. *ECCV*.
4. Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y.M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint*.
5. Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single Shot MultiBox Detector. *ECCV*.
6. **Girshick, R. (2015).** Fast R-CNN. *ICCV.*
7. **Ren, S., He, K., Girshick, R., & Sun, J. (2015).** Faster R-CNN**:** Towards Real**-**Time Object Detection with Region Proposal Networks. *NeurIPS.*
8. **Redmon, J., & Farhadi, A. (2017).** YOLO9000**:** Better**,** Faster**,** Stronger**.** *CVPR.*
9. **Redmon, J., & Farhadi, A. (2018).** YOLOv3: An Incremental Improvement. *arXiv preprint.*
10. **Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y.M. (2022).** YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *arXiv preprint.*
11. **Tan, M., Pang, R., & Le, Q. V. (2020).** EfficientDet: Scalable and Efficient Object Detection. *CVPR.*