

REAL VAQT TASVIR TASNIFI VA OPTIMALLASHTIRISH TEXNIKALARI**Davronov Shoxjahon Rizamat o'g'li****Qarshi davlat universiteti "Algoritmlar va dasturlash texnologiyalari"****kafedrasi mudiri, dotsent, t.f.f.d. (PhD)****Shomamatova Zulxumor Pirmamatovna****Axborot texnologiyalari va menejmenti universiteti magistranti****<https://doi.org/10.5281/zenodo.17627707>****ANNOTATSIYA**

Ushbu tezis real vaqt tasvir tasnifi texnologiyalari, chuqur o'rganishga asoslangan modellar va ularni amaliy tizimlarda samarali ishlatish uchun qo'llaniladigan optimallashtirish texnikalarini tahlil qiladi. Mobil qurilmalar, video monitoring tizimlari, robototexnika va sanoat avtomatizatsiyasi uchun tasvirlarni yuqori tezlikda va aniqlik bilan tasniflash talab etiladi. Shu sababli, konvolyutsion neyron tarmoqlar, yengil arxitekturalar, model siqish, kvantizatsiya, pruning, distillatsiya, ONNX Runtime, TensorRT kabi texnologiyalarning roli batafsil yoritiladi.

KALIT SO'ZLAR

Real vaqt tasvir tasnifi, CNN, MobileNet, EfficientNet, optimallashtirish, kvantizatsiya.

KIRISH

Zamonaviy kompyuter ko'rish tizimlarida tasvir tasnifi eng ko'p qo'llaniladigan asosiy vazifalardan biridir. Real vaqt tasvir tasnifi, ayniqsa, xavfsizlik, tibbiyot, avtomatlashtirilgan ishlab chiqarish, avtonom transport vositalari va mobil ilovalarda muhim ahamiyatga ega. Real vaqt rejimida ishlovchi tizimlar nafaqat aniqlik, balki yuqori tezlikni ham talab qiladi.

Chuqur o'rganish modellarining o'sishi tasvir tasnifi sifatini sezilarli darajada oshirgan bo'lsa-da, bunday modellar ko'p hisoblash resurslarini talab qiladi. Shu sababli, yengil arxitekturalar va modelni optimallashtirish texnikalari real vaqt tizimlarida hal qiluvchi omil bo'lib bormoqda.

ASOSIY QISM

Real vaqt tasvir tasnifining nazariy asoslari

Tasvir tasnifi — tasvirga oid asosiy obyekt yoki holatni aniqlash jarayonidir. Real vaqt tizimlarida asosiy maqsad tasvirni millisekundlar ichida to'g'ri tasniflashdir. Bu jarayonning asosida CNN arxitekturasini, optimallashtirilgan parametrlar, inference tezligini oshirish texnikalari yotadi.

Mashhur yengil tasniflash modellarining arxitekturasini

Real vaqt ish faoliyati uchun CNN modellarining yengil variantlari ishlab chiqilgan. MobileNet oilasi (MobileNetV1–V3) depthwise-separable convolutions texnikasidan foydalanib tezlikni oshiradi va model og'irligini kamaytiradi. EfficientNet esa compound-scaling yondashuvi orqali aniqlik va tezlik o'rtasidagi balansni ta'minlaydi. ShuffleNet va SqueezeNet kabi modellar esa minimal parametrlar bilan yuqori samaradorlik berishga mo'ljallangan.

Optimallashtirish texnikalari

Real vaqt tasnifi uchun modelni yengillashtirish va inference vaqtini qisqartirish muhimdir. Kvantizatsiya model og'irligini 32-bitdan 8-bitga tushiradi va hisoblash tezligini oshiradi. Pruning texnikasi keraksiz neyronlar yoki qatlamlarni o'chirib tashlaydi, bu esa model tuzilmasini ixchamlashtiradi. Distillatsiya orqali katta modeldan kichik modelga bilim o'tkazilib, undan yengil, ammo samarali model olinadi.

Aparatga bog'liq optimallashtirish

GPU, TPU, NPU va boshqa maxsus akseleratorlar real vaqt tasniflash tizimlarining muhim qismidir. TensorRT NVIDIA GPUlari uchun chuqur o'rganish modellarini optimallashtiradi. ONNX Runtime esa turli platformalarda bir xil modelni ishga tushirish imkonini beradi. TFLite mobil qurilmalarda yengil inferensni ta'minlaydi. OpenVINO esa Intel qurilmalari uchun optimallashtirish imkonini beradi.

Amaliy qo'llanilish

Real vaqt tasvir tasnifi videokuzatuv tizimlarida odam yoki obyekt holatini aniqlashda qo'llaniladi. Avtonom robotlar kameradan kelayotgan tasvirlarni darhol tahlil qilib, harakat yo'nalishini aniqlaydi. Mobil ilovalarda esa tasvir tasnifi yordamida mahsulotlarni aniqlash, botanika ilovalarida o'simliklarni tanish yoki tibbiy skanerlardan olingan tasvirlarni tezkor tasniflash mumkin.

XULOSA

Tezida zamonaviy real vaqt tasvir tasnifi texnologiyalari va optimallashtirish usullari keng yoritildi. Chuqur o'rganishga asoslangan modellar yuqori aniqlikni ta'minlagani uchun amaliyotda juda samarali bo'lsada, ularning resurs talabchanligi optimallashtirish texnikalarini qo'llashni talab qiladi. Yengil modellar – MobileNet, EfficientNet va ShuffleNet real vaqt uchun juda qulay bo'lsa, TensorRT, TFLite va ONNX Runtime kabi akseleratsiya texnologiyalari inference tezligini yanada oshiradi. Kelgusida transformerga asoslangan tasniflash modellarining yengil versiyalari va edge computing uchun samarador modellar rivojlanishi kutilmoqda.

Foydalanilgan adabiyotlar:

1. Howard, A. et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
2. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
3. Han, S. et al. (2016). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding.
4. NVIDIA Corporation. (2022). TensorRT Developer Guide.
5. Google Research Team. (2020). TensorFlow Lite Model Optimization Toolkit.
6. Microsoft. (2021). ONNX Runtime: High-Performance Machine Learning Inference.