Računarski fakultet

DIPLOMSKI RAD

Slika je vredna 16x16 reči: Vision Transformeri

Vanja Kovinić
RN 42/2020

Mentor:
dr Nemanja Ilić

Komisija:
dr Nemanja Ilić
dr Nevena Marić

Beograd, septembar 2024.

Apstrakt

Ovaj diplomski rad istražuje **Vision Transformere** (**ViT**), nov pristup u oblasti računarskog vida koji koristi arhitekturu transformera prvobitno razvijenu za obradu prirodnog jezika. Prvi deo rada pruža detaljan pregled arhitekture transformera, uključujući ključne komponente kao što su ***self-attention*** **mehanizam** i **poziciono enkodovanje**, i diskutuje njihove svrhe i funkcionalnosti. Nakon toga, fokus se prebacuje na Vision Transformere, objašnjavajući kako se slike transformišu u **tokene** i obrađuju kroz **enkoder transformera** kako bi se primenili na rešavanje vizuelnih zadataka.

Rad zatim ulazi u praktične aspekte implementacije Vision Transformera, uključujući izbor i podešavanje **hiperparametara** za poboljšanje performansi. Izvršeno je i poređenje sa referentnim implementacijama, i predložen pristup za poboljšanje performansi. Prikazani su različiti eksperimenti, zajedno sa diskusijom njihovih rezultata, pružajući uvid u efikasnost i izazove povezane sa Vision Transformerima.

Na kraju, rad naglašava značaj Vision Transformera u oblasti računarskog vida, prikazujući njihov potencijal i ograničenja, kao i njihove praktične primene.

1 Uvod

"Pre otprilike 540 miliona godina, Zemlja je bila obavijena tamom. Ovo nije bilo zbog nedostatka svetlosti, već zato što organizmi još uvek nisu razvili sposobnost da vide. Iako je sunčeva svetlost mogla da prodre u okeane do dubine od 1.000 metara i hidrotermalni izvori na dnu mora isijavali svetlost u kojoj je život cvetao, nijedno oko nije se moglo naći u tim drevnim okeanima, nijedna retina, rožnjača ili sočivo. Sva svetlost i život nikada nisu viđeni. Koncept gledanja nije ni postojao tada i ova sposobnost nije ostvarena sve dok nije stvorena.

Iz nama nepoznatih razloga, trilobiti su se pojavili kao prva bića sposobna da spoznaju svetlost. Oni su prvi prepoznali da postoji nešto izvan njih samih, svet okružen višestrukim jedinkama. Rađenje vida se smatra da je pokrenulo kambrijsku eksploziju, period u kojem se veliki broj vrsta životinja pojavljuje u fosilnom zapisu. Vid je započeo kao pasivno iskustvo, jednostavno propuštanje svetlosti, ali je ubrzo postao aktivniji. Nervni sistem je počeo da evoluira, vid je prešao u uvid, gledanje je postalo razumevanje, a razumevanje je dovelo do akcije, a sve to je dovelo do nastanka inteligencije.

Danas nismo više zadovoljni vizuelnom spoznajom koju nam je priroda dala. Radoznalost nas je navela da stvorimo mašine koje mogu da "vide" kao mi, pa čak i inteligentnije." - Li Fei-Fei [1]

1.1 Istorija i motivacija

Some random text.

References

[1] Li Fei-Fei. *With spatial intelligence, AI will understand the real world.* URL: https://www.ted.com/talks/fei_fei_li_with_spatial_intelligence_ai_will_understand_the_real_world.