



**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.