Generativne glave na CLIP značajkama za klasifikaciju s malo primjera

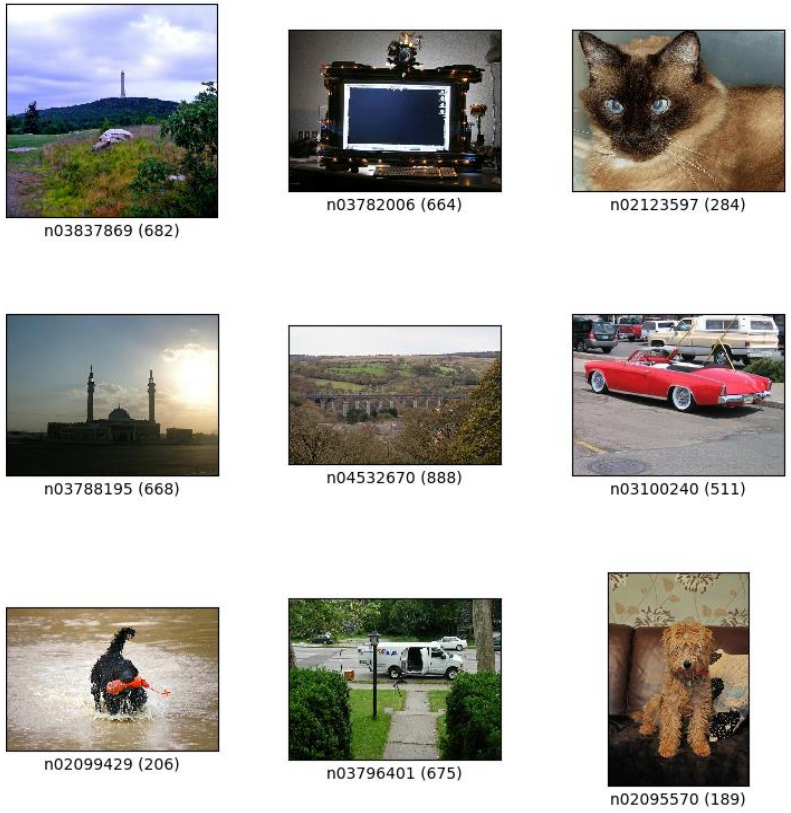
Dominik Agejev, Ivan Skukan, Ian Marković,
Lucija Petkoviček, Leonora Đemaili

GitHub 

Motivacija

- Kako pouzdano klasificirati primjere koji su izvan distribucije (OOD primjeri)?
- Primjer:
  - AI sustav treniran na 1000 bolesti, a pacijent ima rijetku bolest
  - Hoće li sustav reći "ne znam" ili dati pogrešnu dijagnozu?
- Primjer:
  - AI sustav treniran za autonomnu vožnju u gradskim uvjetima
  - Slon pobjegne iz zoološkog. Što će sustav zaključiti kad ga vidi?

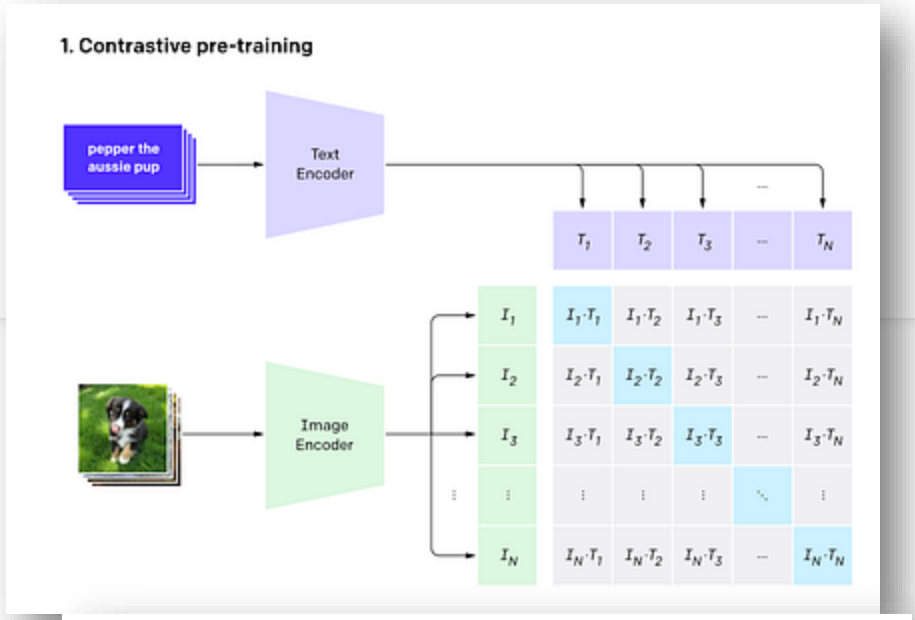
Definicija problema

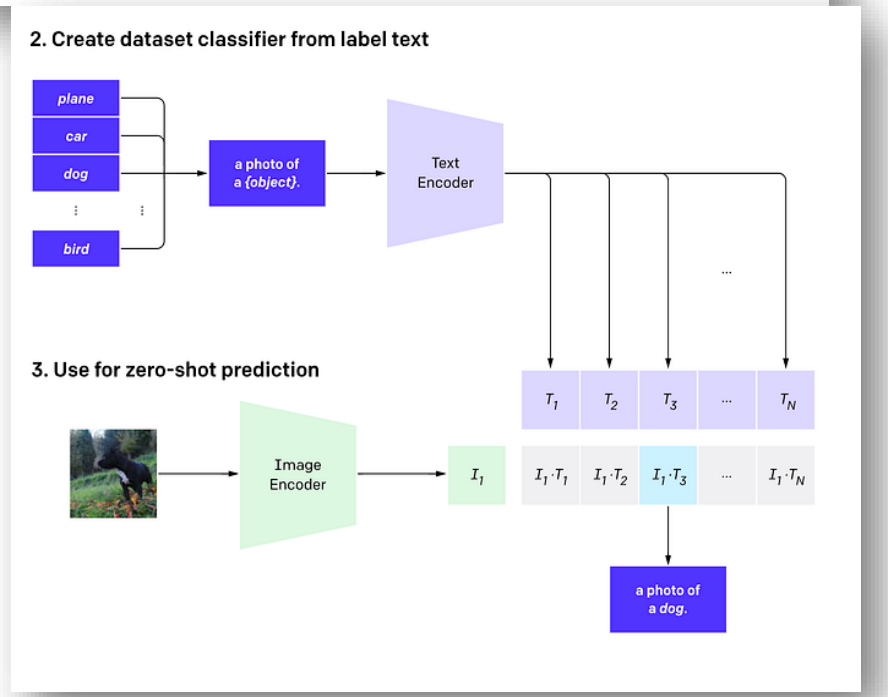


- CLIP određuje koliko je slika slična tekstu, ali ne povlači granicu između poznatih (ID) i nepoznatih (OOD) klasa
- Zato treniramo klasifikacijske glave koje treniramo na zamrznutom CLIP-u
- ImageNet-1k za podatke iz distribucije
- ImageNet-O za podatke van distribucije (OOD)
- K-shots (broj označenih primjera klase za treniranje glave): 0,1,2,4,8,16

Što je CLIP?

- Contrastive Language-Image Pretraining
- Model povezuje slike i tekst u zajednički semantički prostor:
  - "A photo of a dog" -> blizu slike psa u prostoru ugrađivanja
- Arhitektura:
  - Encoder za sliku i encoder za tekst
  - Učenje na temelju kontrasta vektora slike i teksta
- Zašto CLIP?
  - Predtreniran na 400m slika-tekst parova
  - Zero-shot - može klasificirati s 0 primjera (baseline daljnje analize)





Klasifikacijske glave

- Zero-shot:
  - Ugrađeni tekst: "A photo of {class}" za svih 1000 klasa
  - Uzimamo prosjek više promptova
  - Ne treniramo
- Linearna glava:
  - Regularizirani linearni model klasificira ugrađivanja

- **Prototipna glava:**
  - Temelji se na kosinusnoj udaljenosti
  - Klasifikacija: najbliži prototip
  - OOD: daleko od svih prototipa
- **Gaussova glava:**
  - Za svaku klasu učimo parametre
  - Pretpostavljamo zajedničku kovarijacijsku matricu
  - Problem: 512x512 parametara

Diskriminativne **Generativne**

Ključna pitanja

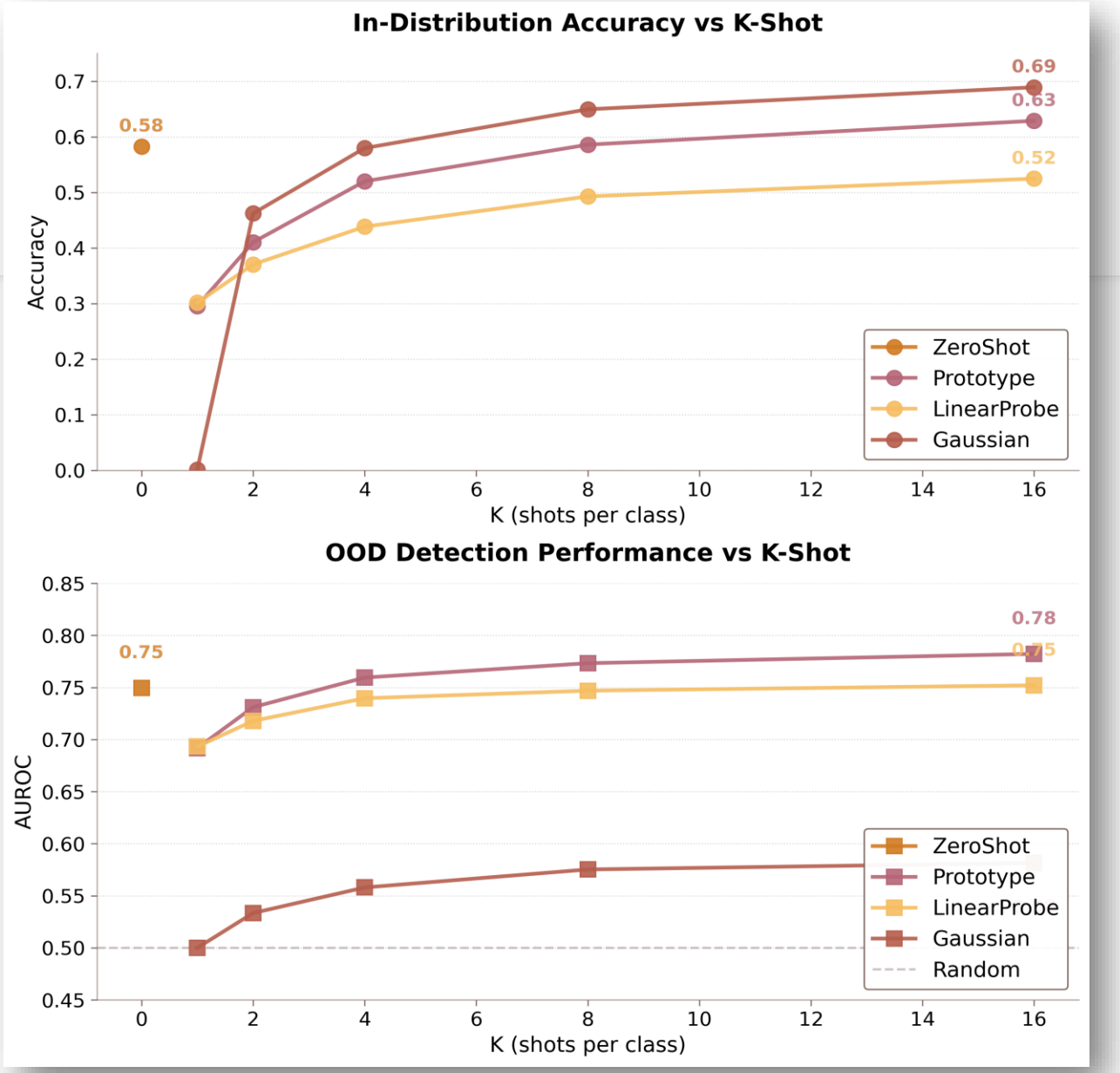
Koja metoda najbolje balansira ID točnost i OOD detekciju?

Koliko primjera je potrebno da nadmašimo zero-shot?

Ako neka metoda ne uspijeva, zašto?

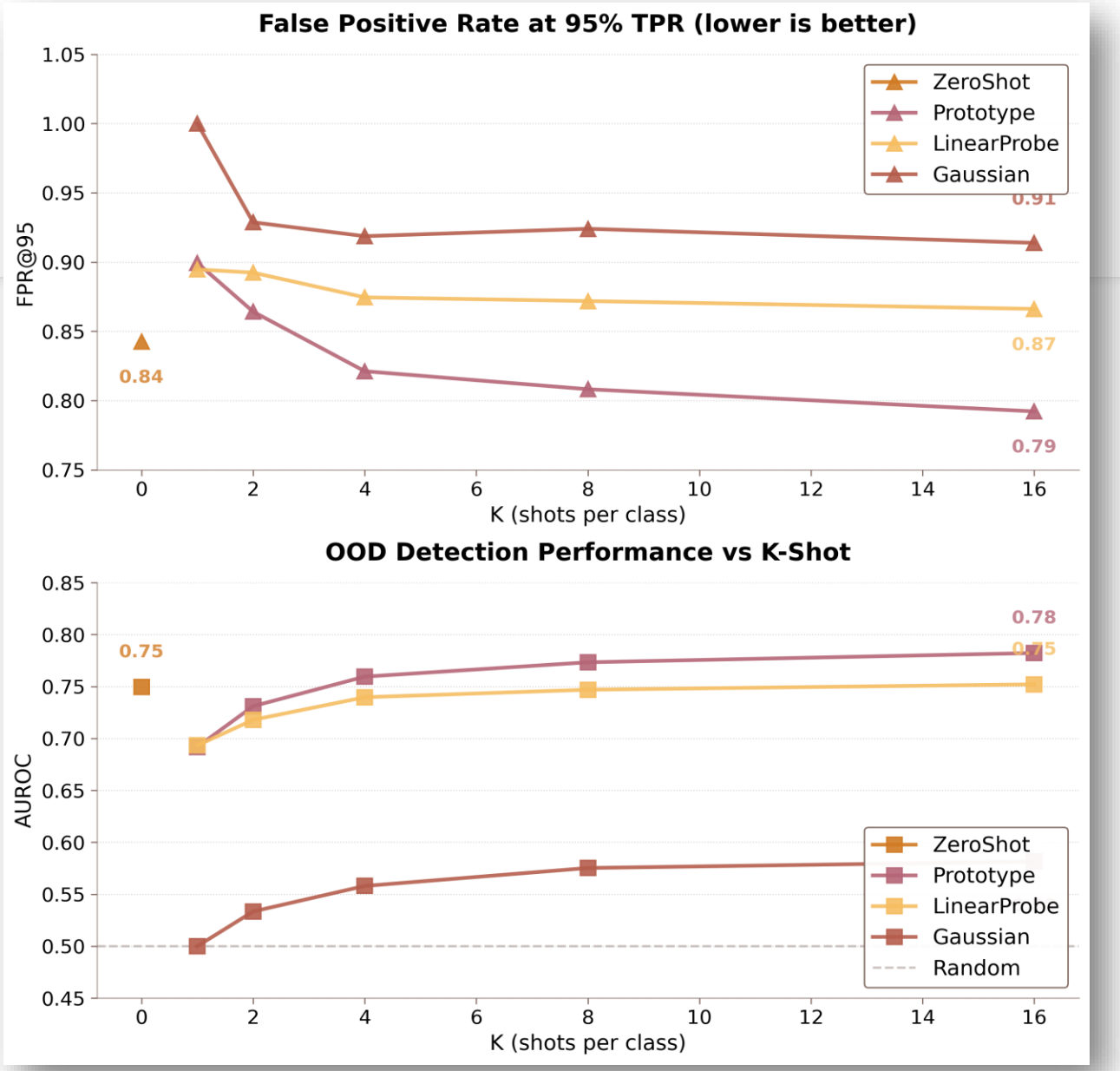
Pregled rezultata

- Zero-shot je na 58.3%
- Gaussian neuspješan za OOD
- Prototype je najbolji za OOD AUROC
- Prototype je najbolji za FPR@95 (Niže je bolje)
- Prototype ima najbolji trade-off



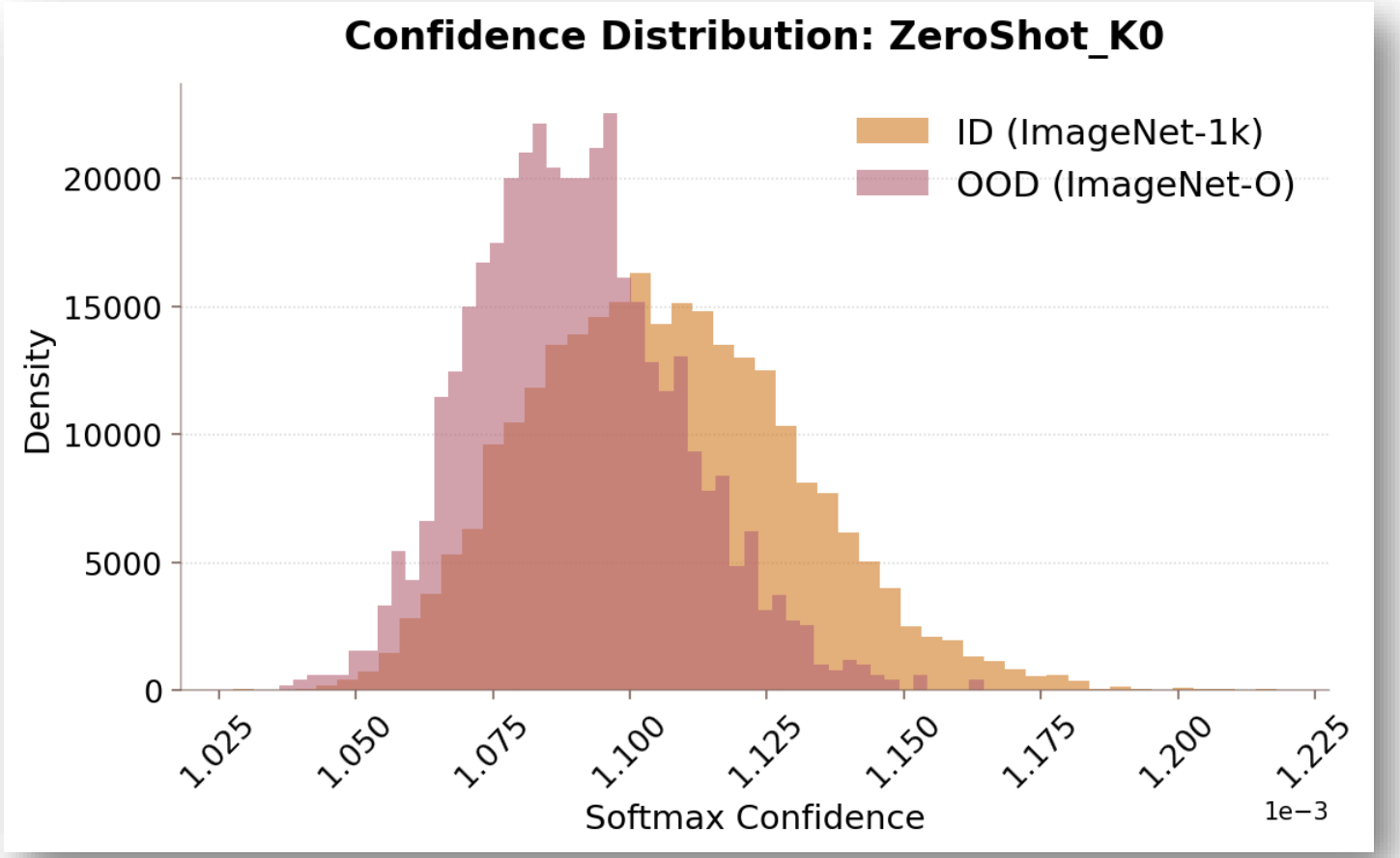
Pregled rezultata

- Zero-shot je na 58.3%
- Gaussian neuspješan za OOD
- Prototype je najbolji za OOD AUROC
- Prototype je najbolji za FPR@95 (Niže je bolje)
- Prototype ima najbolji trade-off



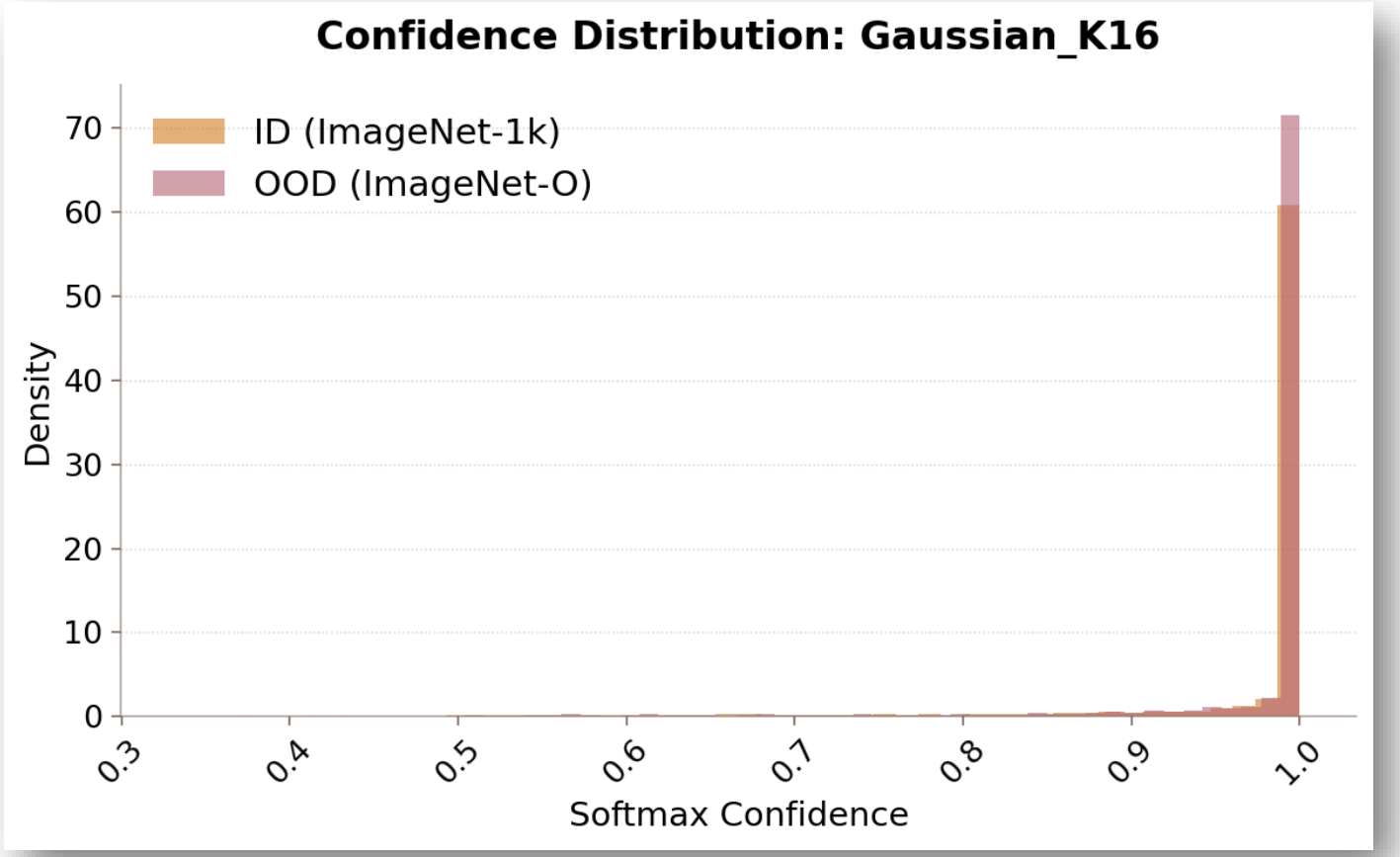
Pregled rezultata – zero-shot

- Zero-Shot (K=0): 58.3% accuracy
- Few-Shot (K=4): 52.0% accuracy (Prototype)
- Idealno bi distribucije bile potpuno odvojene
- Slaba sigurnost!



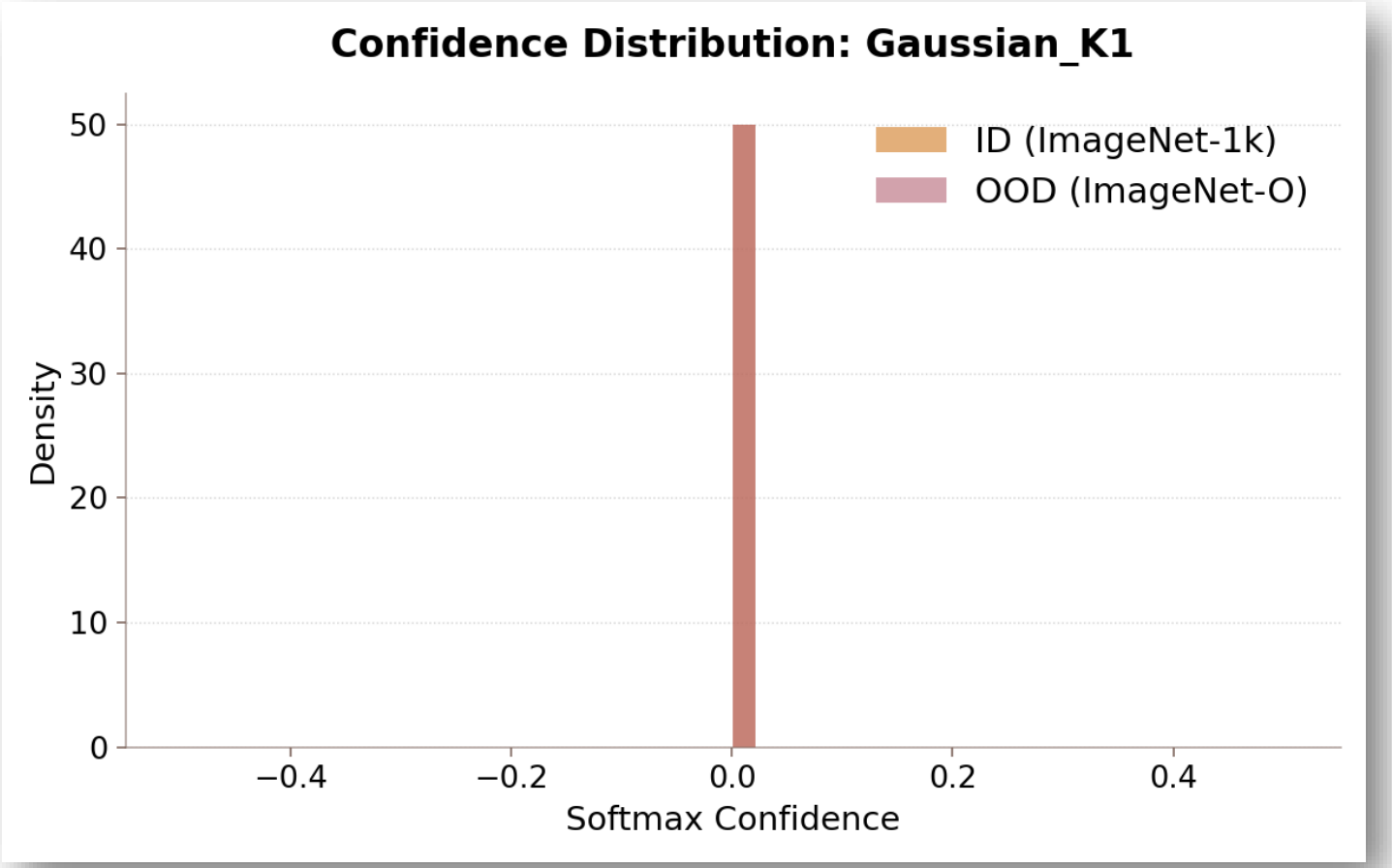
Pregled rezultata – Gaussian

- Gaussian daje najveću ID točnost:
  - 16 uzoraka po klasi daje dovoljno informacije za procjenu distribucije
- Model je vrlo siguran u svoje (često krive) predikcije za K = 16.
- Skoro pa ne razdvaja distribucije
- Za K=1 ne razdvaja i ima slabu sigurnost (ne zna što se događa)



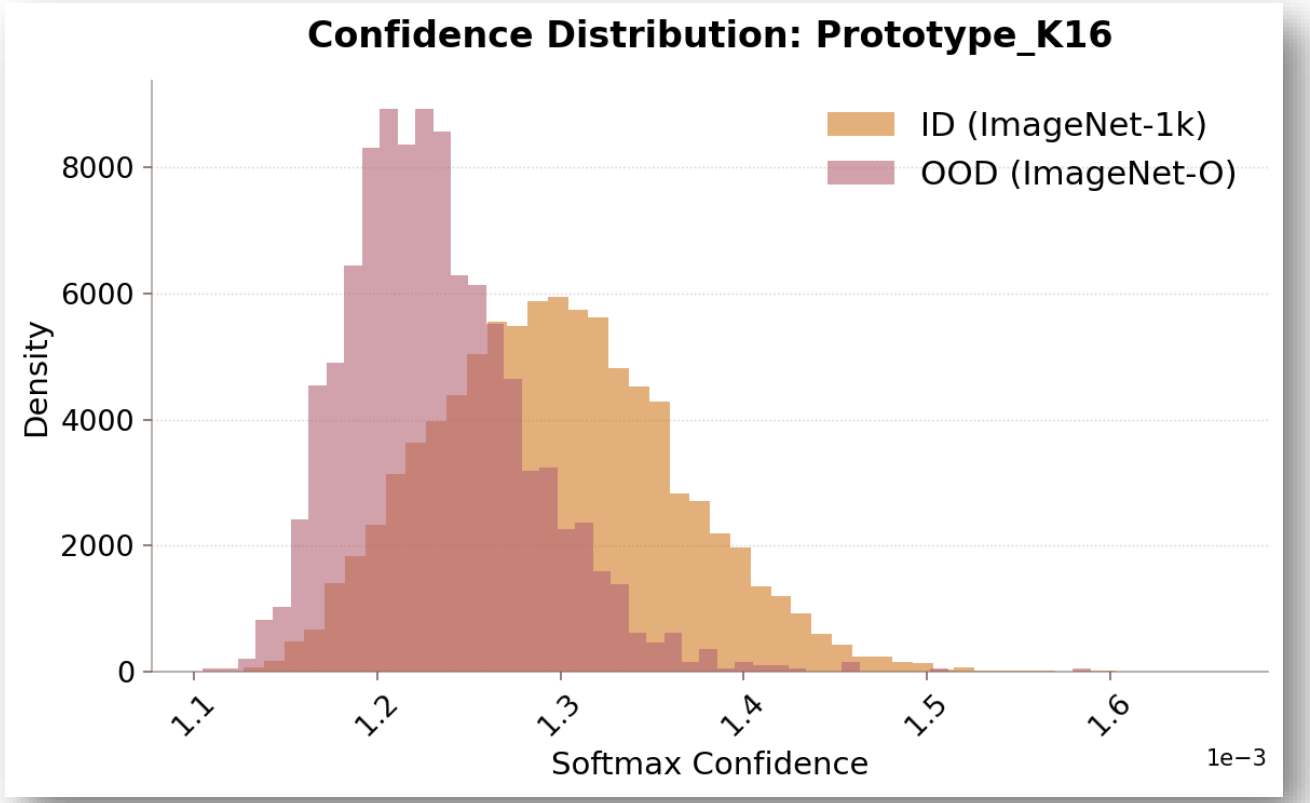
Pregled rezultata – Gaussian

- Gaussian daje najveću ID točnost:
  - 16 uzoraka po klasi daje dovoljno informacije za procjenu distribucije
- Model je vrlo siguran u svoje (često krive) predikcije za K = 16.
- Skoro pa ne razdvaja distribucije
- Za K=1 ne razdvaja i ima slabu sigurnost (ne zna što se događa)



Pregled rezultata - Prototip (K=16)

- ID točnost je 62.9%
- Marginalna ali konzistentna poboljšanja
- Bolja OOD detekcija – distribucije malo udaljenije
- Još uvijek je veliko preklapanje i mala sigurnost!



Sažetak

- Prototype glava jedina nadmašuje zero-shot uz K>=8
- Slaba sigurnost vjerojatno je uzrokovana velikim brojem klasa (ImageNet je težak)
- Gaussian:
  - K=1 – Prokletstvo dimenzionalnosti (262k parametara)
  - K>1 – Jaka samouvjerenost - misli da je sve sigurno ImageNet-1k
- Linear probe:
  - Lošiji rezultati od zero-shot
  - Unakrsna entropija optimira točnost, a ne sigurnost

Hvala na pažnji!

- Pitanja?

Izvori

- CLIP slike: https://medium.com/@paluchasz/understanding-openais-clip-model-6b52bade3fa3
- ImageNet slika: https://www.tensorflow.org/datasets/catalog/imagenet2012_subset