Izvještaj

Motivacija

AI sistemi često zahtijevaju velike količine podataka kako bi postigli visoke performanse. Međutim, u mnogim realnim scenarijima jednostavno nema dovoljno podataka za adekvatno treniranje i evaluaciju.

Stvarni scenariji:

- **Medicinska dijagnostika**: Ako model prepoznaje 1000 bolesti, a pacijent dođe sa 1001. koju model nikad nije vidio, cilj je da sistem kaže "ne znam" umjesto postavljanja pogrešne dijagnoze s visokom sigurnošću1.
- **Autonomna vožnja**: U nepredviđenim situacijama (npr. slon na cesti), važno je da model prepozna nepoznat objekt i reaguje oprezno radi sigurnosti putnika2.
- **Industrijska kontrola kvaliteta**: Sistem mora detektovati nove defekte na proizvodima umjesto da ih klasifikuje u već poznate kategorije3.

Ključni izazovi:

1. **Few-Shot Learning**: Kako naučiti robusne modele sa svega 1–16 uzoraka po klasi? 4
2. **Out-of-Distribution (OOD) Detection**: Kako prepoznati uzorke koji ne pripadaju nijednoj treniranoj klasi? 5

Temeljni kompromis: Potrebno je balansirati između točnosti na poznatim klasama (ID accuracy) i opreza prema nepoznatim uzorcima (OOD detection), jer modeli sigurni na ID podacima često pokazuju preveliku sigurnost i na OOD podacima6.

Definicija problema

Zadatak: Few-shot klasifikacija uz detekciju out-of-distribution uzoraka7.

Setup:

- **In-Distribution (ID)**: ImageNet-Val dataset (1000 klasa, 50.000 slika)8.
- **Out-of-Distribution (OOD)**: ImageNet-O dataset (2000 slika, objekti van ImageNet taksonomije)9.
- **Few-Shot režim**: $K \in \{0, 1, 2, 4, 8, 16\}$ uzoraka po klasi10.

Arhitektura:

- **Feature Extraction**: CLIP ViT-B-16 (pre-trained, 512-dimenzionalni embedding)11.
- **Classification Heads**:

1. **Zero-Shot**: Kosinusna sličnost s tekstualnim embeddingom12.
2. **Prototype**: Udaljenost do centroida klase13.
3. **Linear Probe**: Linearna transformacija (512 → 1000)14.
4. **Gaussian**: Gaussova diskriminativna analiza s Mahalanobisovom udaljenosti15.

Metrike:

- **ID**: Accuracy (točnost) i ECE (Expected Calibration Error)16.
- **OOD**: AUROC (Area Under ROC) i FPR@95 (False Positive Rate pri 95% TPR)17.

---

1. Detaljna analiza rezultata

1.1 Neuspjeh Gaussian Heada

Gaussian head potpuno zakazuje pri **K=1** (točnost svega **0.11%**), dok pri **K=16** postiže visoku točnost (68.9%), ali lošu OOD detekciju (AUROC=0.571)18.

Uzroci:

- **Prokletstvo dimenzionalnosti**: 262.144 parametara kovarijanse naspram samo jednog uzorka pri K=1 čini matricu singularnom19.
- **Kršenje pretpostavke**: Klase u CLIP prostoru ne dijele istu strukturu kovarijanse (neke su uske, neke široke)20.
- **Overconfidence**: Model prelazi iz potpune nesigurnosti u ekstremnu sigurnost bez "sredine", čineći ID i OOD distribucije identičnim21.

1.2 Točnost vs. OOD detekcija

Metoda	K=16 Točnost	K=16 AUROC	K=16 FPR@95
Linear Probe	52.5%	0.752	86.6%
Prototype	62.9%	0.782	79.3%
Gaussian	68.9%	0.571	91.5%

- **Linear Probe** je uspješan pri malom K zbog diskriminativnog treniranja, ali je previše samouvjeren na OOD podacima22.
- **Prototype** dominira pri K≥8 jer čuva CLIP semantiku i bolje razdvaja ID od OOD distribucija23.

1.3 Kalibracija (ECE)

Prosječni ECE je 0.413, što ukazuje na lošu kalibraciju24. Pri K=16, Gaussian je najbolje kalibriran (0.299), dok je Prototype najlošiji (0.628), što znači da Prototype zahtijeva

dodatni _temperature tuning_25.

1.4 Zero-Shot Baseline

Zero-shot CLIP (Accuracy: 58.3%, AUROC: 0.750) nadmašuje sve metode dok se ne dosegne **K≥8** uzoraka26.

---

2. Konačni sažetak rezultata

Performanse po metodama (prosjek):

K-shot	Metoda	Accuracy	AUROC	FPR@95	ECE
0	ZeroShot	58.3%	0.750	84.3%	0.582
1	LinearProbe	30.2%	0.693	89.5%	0.301
1	Prototype	29.6%	0.692	89.9%	0.294
1	Gaussian	0.1%	0.500	100%	0.000
4	Prototype	52.0%	0.760	82.2%	0.519
16	Prototype	62.9%	0.782	79.3%	0.628
16	Gaussian	68.9%	0.571	91.5%	0.299

Zaključci:

1. **Gaussian Head** je neupotrebljiv pri K=1 zbog visoke dimenzionalnosti27.
2. **Zero-shot CLIP** je najbolji izbor za vrlo mali broj podataka (K < 8)28.
3. **Prototype Head pri K=16** nudi najbolji balans između točnosti i sigurnosti detekcije29.
4. **Distance-based metode** u CLIP semantičkom prostoru nadmašuju diskriminativne metode u few-shot OOD zadacima30.

Ključna poruka:

> "Prototype head sa K=16 uzoraka postiže najbolju ravnotežu između ID točnosti (62.9%) i OOD detekcije (AUROC=0.782), nadmašujući zero-shot CLIP za 4.6% u točnosti." 31