ROBUST FINE-TUNING OF ZERO-SHOT MODELS

Докладчик: Абрамов Арсений
Рецензент-исследователь: Клименко Злата
Хакер: Присяжнюк Артём

14 декабря, 2022

Введение

ImageNet (Deng et al.)

оригинал

ObjectNet (Barbu et al.)

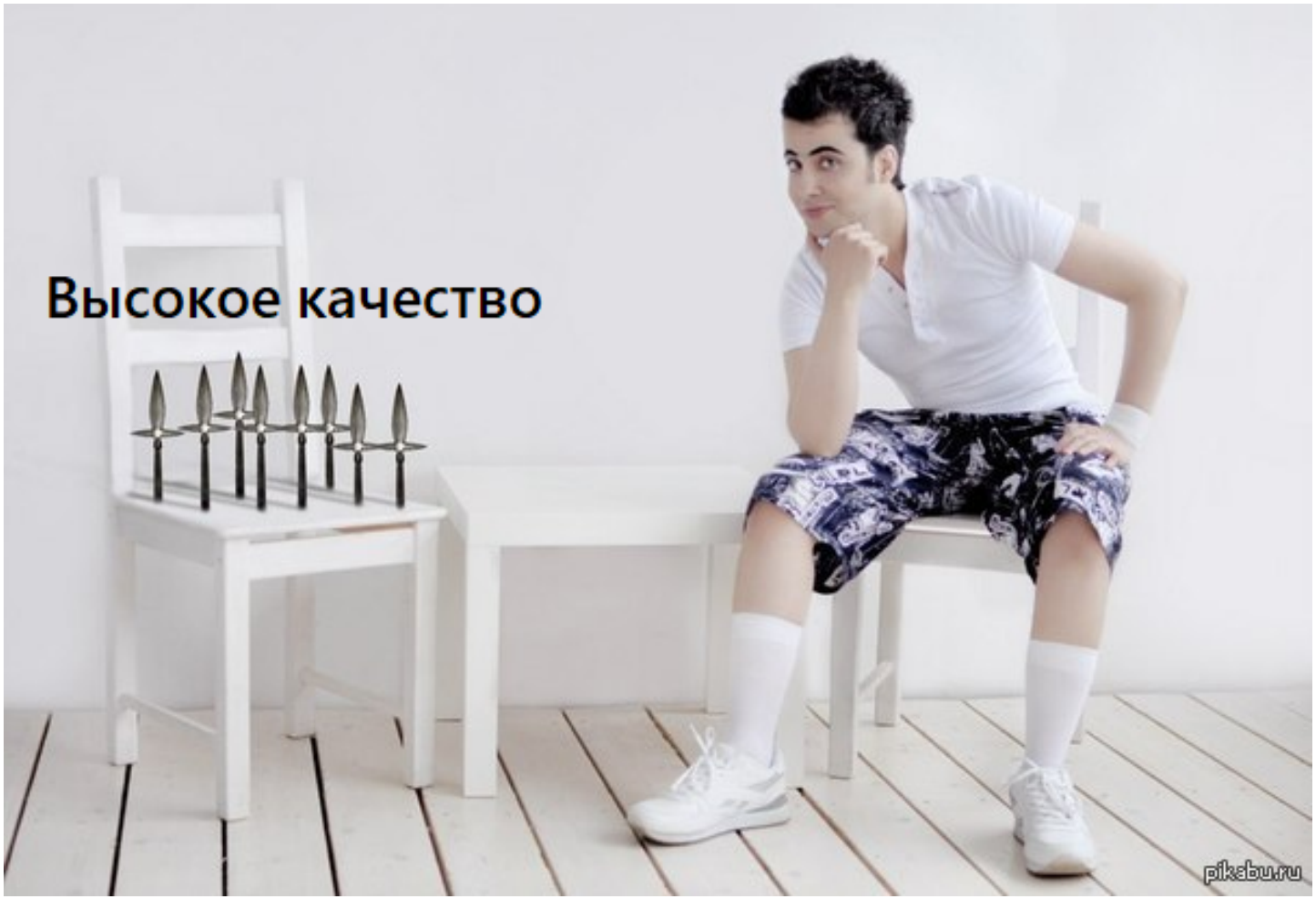
смещённое

Введение

Zero-shot: предобученная модель; в тестовой выборке используются классы, не задействованные на обучающей выборке.

Robustness: насколько меняется качество модели при смене распределения.

он выбрал Robustness



Введение

Zero-shot:

- Robustness ✓
- Accuracy ✓

Fine tune:

- Robustness ✓
- Accuracy ✓

Введение

Как дообучать без потери обобщающей способности?

WiSE-FT

WiSE-FT := Weight-Space Ensembles for Fine-Tuning

WiSE-FT

WiSE-FT:

- Robustness ✓
- Accuracy ✓

WiSE-FT

Первый шаг:
Fine-tuning zero-shot модели.

Второй шаг:
Линейная интерполяция между весами оригинальной и дообученной моделей (Weight-Space ensembling)

Псевдокод

Algorithm 1 Pytorch pseudocode for WiSE-FT

```
def wse(model, zeroshot_checkpoint, finetuned_checkpoint, alpha):
    # load state dicts from checkpoints
    theta_0 = torch.load(zeroshot_checkpoint)["state_dict"]
    theta_1 = torch.load(finetuned_checkpoint)["state_dict"]

    # make sure checkpoints are compatible
    assert set(theta_0.keys()) == set(theta_1.keys())

    # interpolate between all weights in the checkpoints
    theta = {
        key: (1-alpha) * theta_0[key] + alpha * theta_1[key]
        for key in theta_0.keys()
    }

    # update the model (in-place) according to the new weights
    model.load_state_dict(theta)

def wise_ft(model, dataset, zeroshot_checkpoint, alpha, hparams):
    # load the zero-shot weights
    theta_0 = torch.load(zeroshot_checkpoint)["state_dict"]
    model.load_state_dict(theta_0)

    # standard fine-tuning
    finetuned_checkpoint = finetune(model, dataset, hparams)

    # perform weight-space ensembling (in-place)
    wse(model, zeroshot_checkpoint, finetuned_checkpoint, alpha)
```

Главное

- работаем с CLIP;
- ансамблирование вычислительную сложность не увеличивает;
- α (он же *mixing coefficient*) разумно брать за 0.5;
- при дообучении только линейного классификатора

эквивалентно простому ансамблированию;

- linear mode connectivity;

Ээээксперименты

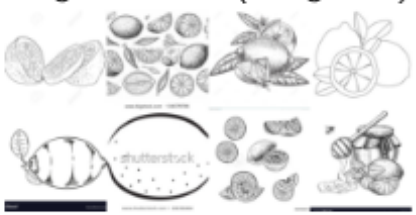
	IN (reference)	Distribution shifts					Avg shifts	Avg ref., shifts
		IN-V2	IN-R	IN-Sketch	ObjectNet*	IN-A		
CLIP `ViT-L/14@336px`								
Zero-shot [82]	76.2	70.1	88.9	60.2	70.0	77.2	73.3	74.8
Fine-tuned LC [82]	85.4	75.9	84.2	57.4	66.2	75.3	71.8	78.6
Zero-shot (PyTorch)	76.6	70.5	89.0	60.9	69.1	77.7	73.4	75.0
Fine-tuned LC (ours)	85.2	75.8	85.3	58.7	67.2	76.1	72.6	78.9
Fine-tuned E2E (ours)	86.2	76.8	79.8	57.9	63.3	65.4	68.6	77.4
WiSE-FT (ours)								
LC, α=0.5	83.7	76.3	89.6	63.0	70.7	79.7	75.9	79.8
LC, optimal α	85.3	76.9	89.8	63.0	70.7	79.7	75.9	80.2
E2E, α=0.5	86.8	**79.5**	89.4	64.7	71.1	79.9	76.9	81.8
E2E, optimal α	**87.1**	**79.5**	**90.3**	**65.0**	**72.1**	**81.0**	**77.4**	**81.9**

ImageNet (Deng et al.)

ImageNetV2 (Recht et al.)

ImageNet-R (Hendrycks et al.)

ImageNet Sketch (Wang et al.)



ObjectNet (Barbu et al.)

ImageNet-A (Hendrycks et al.)

Источники

- https://arxiv.org/abs/2109.01903