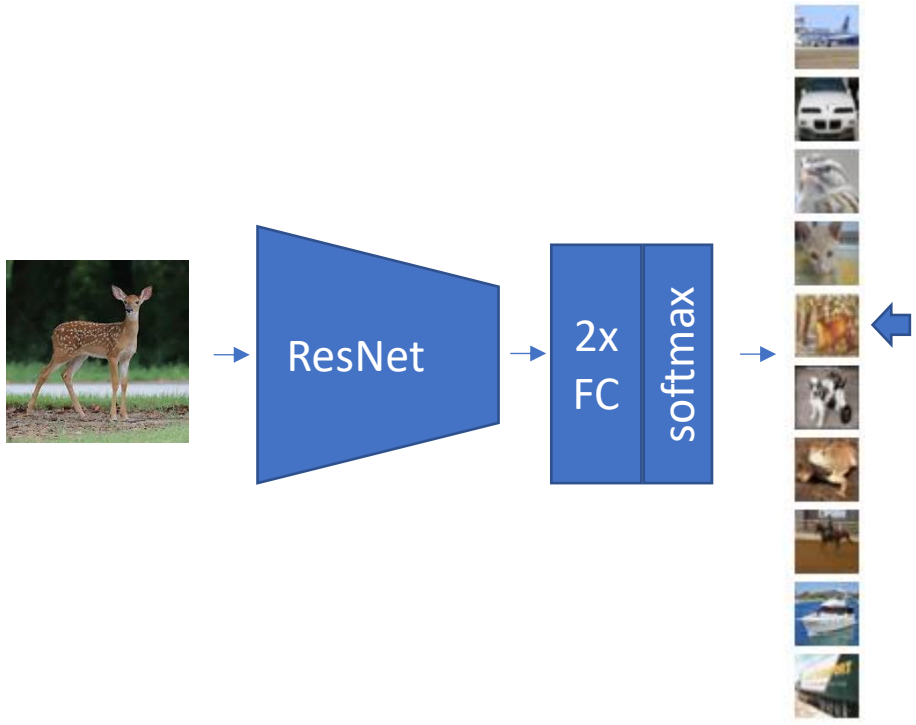
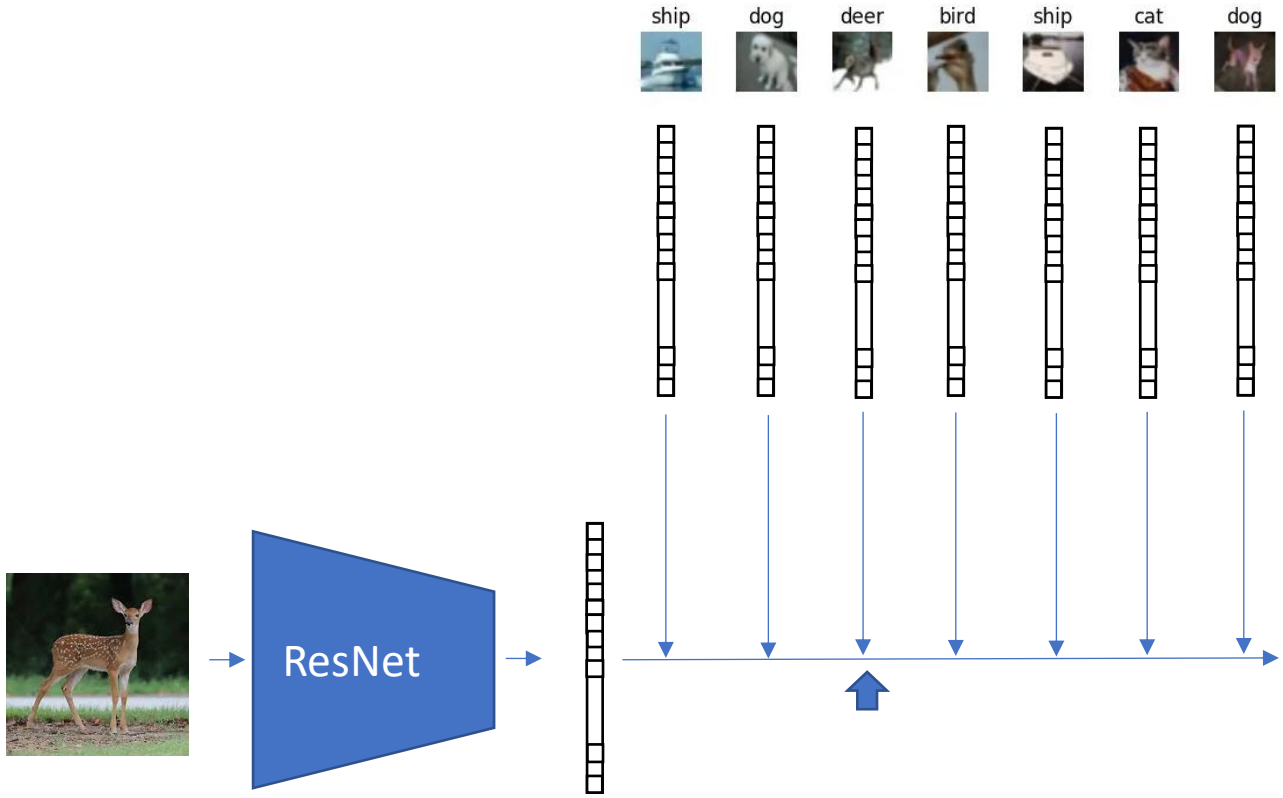


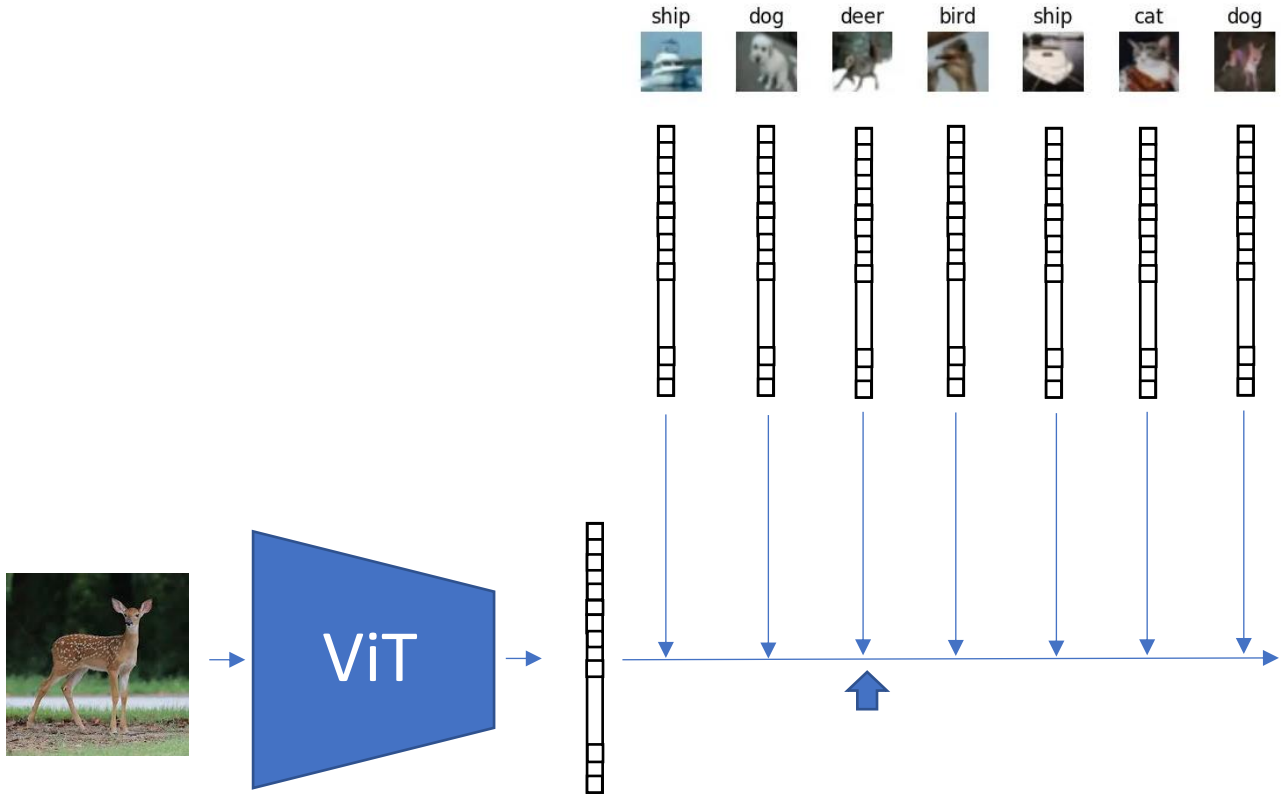
10 pravdepodobností



2048 príznakov

Euclidean distance

Enrollment



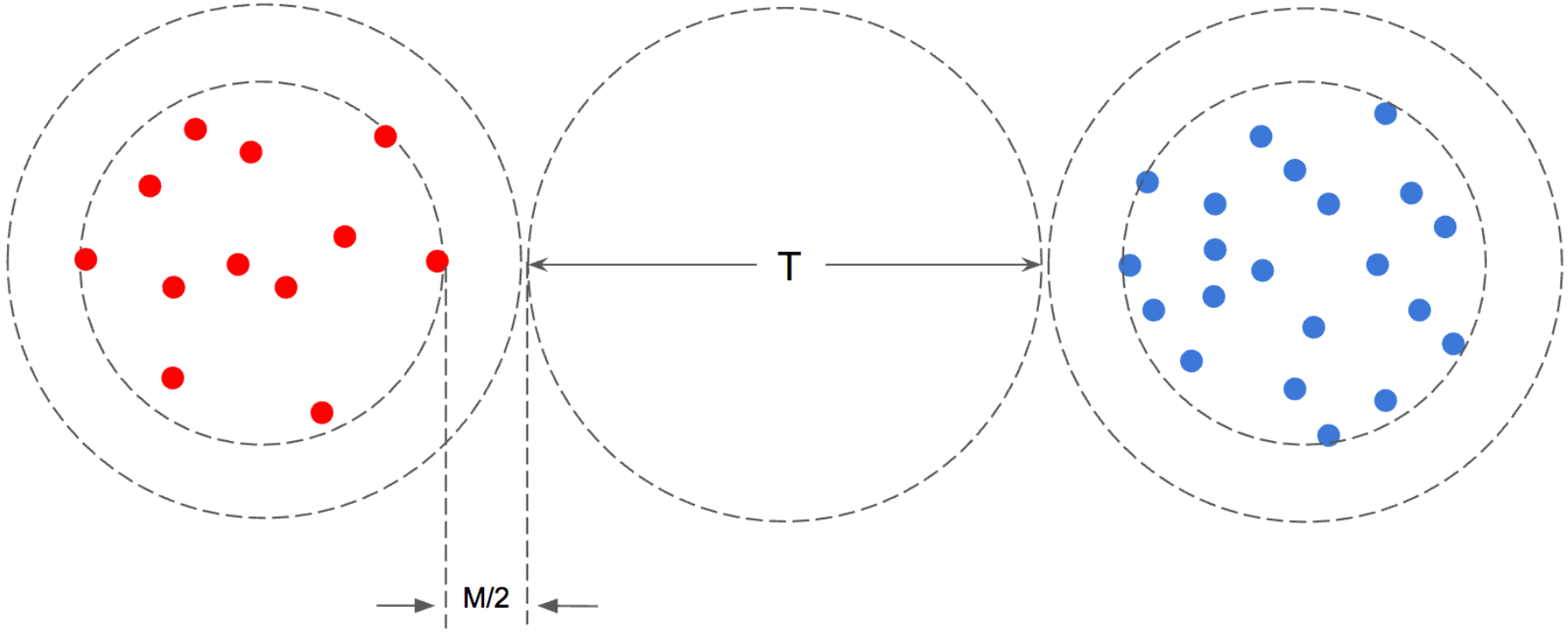
384 príznakov

Scalar product
(cosine similarity)

Enrollment

Contrastive (Metric) Learning

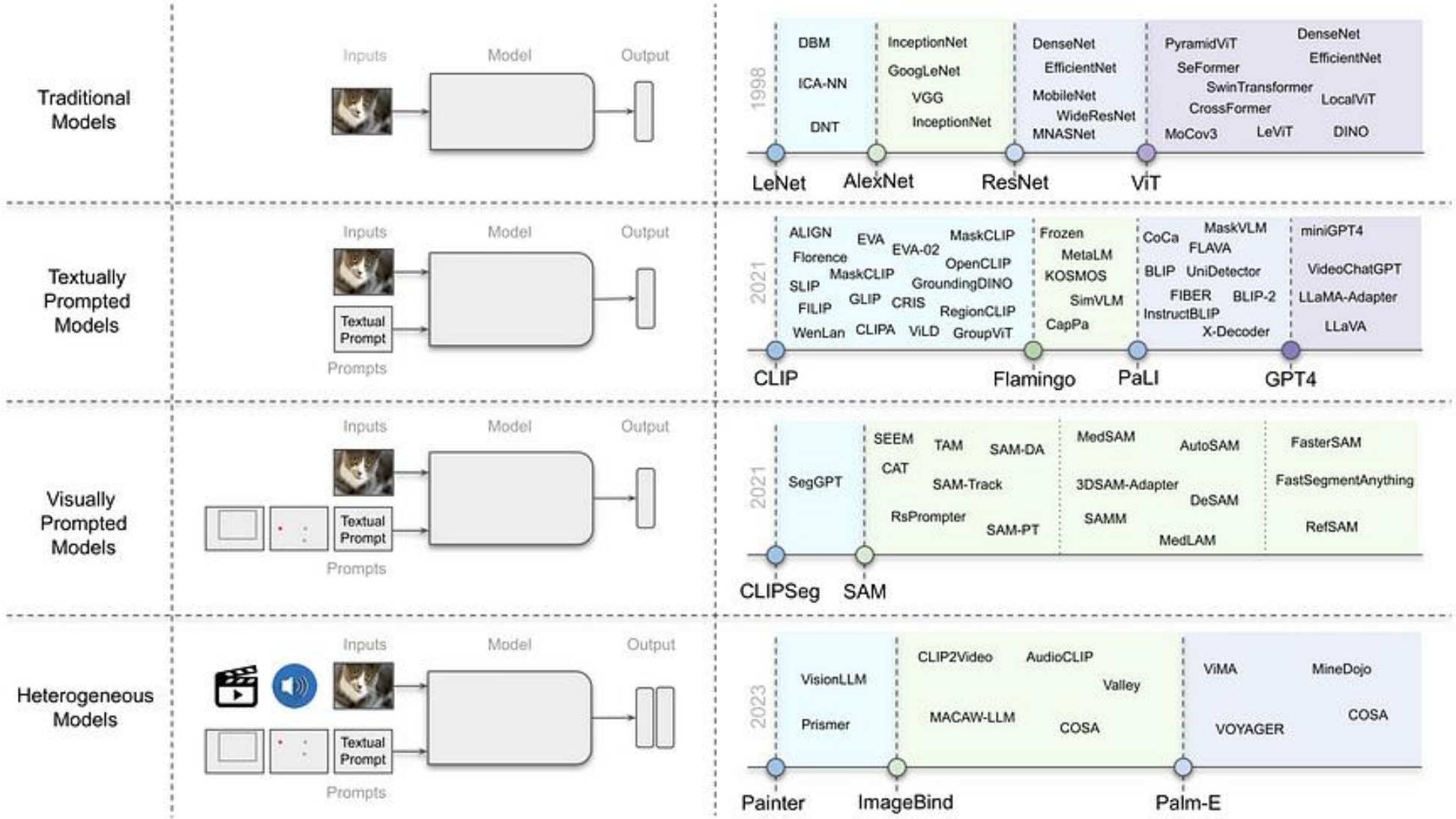
- Chybovú funkciu nedefinujem podľa porovnania výstupu a želaného výstupu, ale podľa toho, že výstupy majú spĺňať určitú metriku



Foundation modely

- modely, ktoré nie sú určené na trénovanie používateľom, ale používajú sa tak ako sú
- Systém nimi vybavený je schopný buď zero-shot, one-shot či few-shot learningu
- Väčšina z nich má prompt a niektoré umožňujú prompt engineering, čím sa realizuje tzv. in-context learning (vylepšuje sa odozva bez zmeny váh, mení sa len momentálny vnútorný stav)
- Sú natrénované z obrovského množstva dát, z veľkej časti spôsobom self-supervision
- Prístup k strojovému učeniu: data-centric → model-centric

Foundation modely



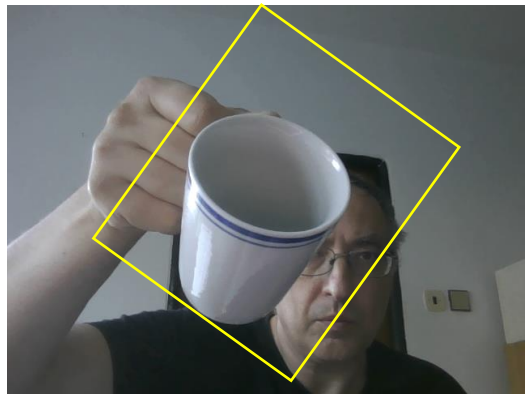
zdroj obrázku: https://arxiv.org/pdf/2307.13721.pdf

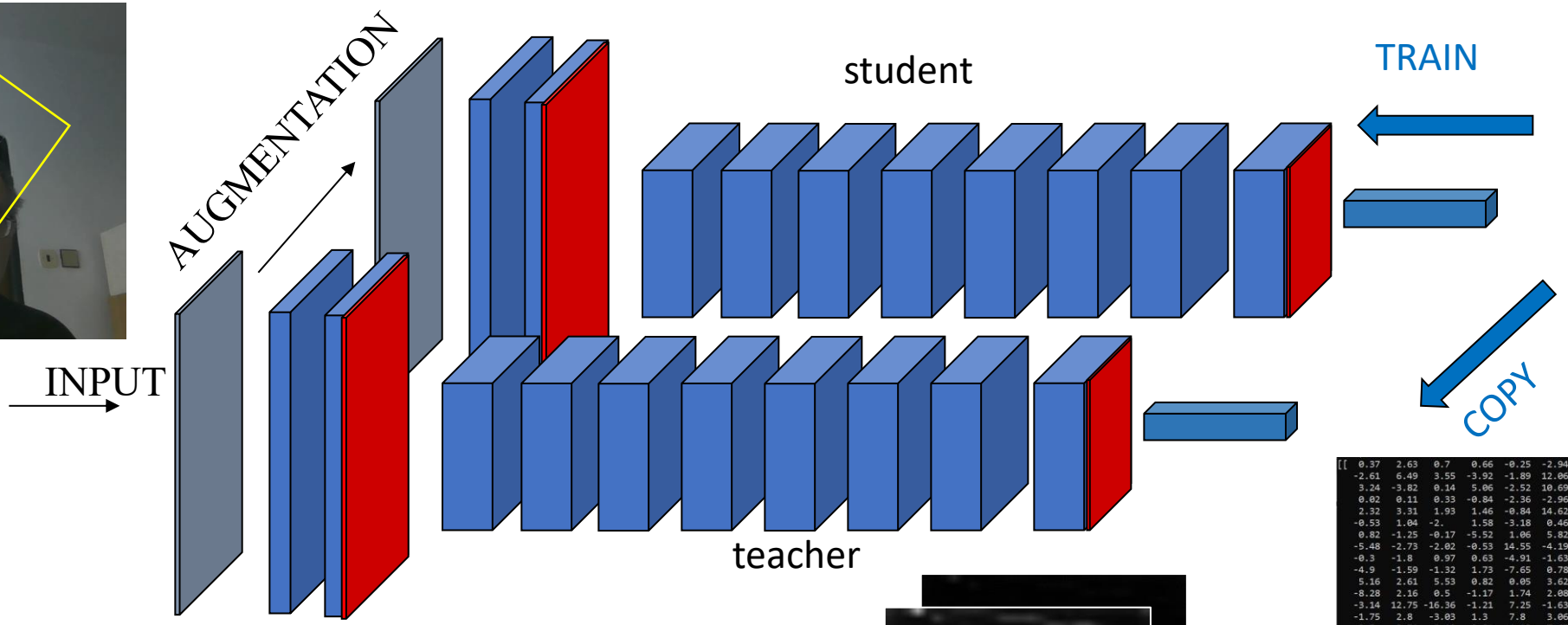
DINO (self-DIstillation with NO Labels) v1

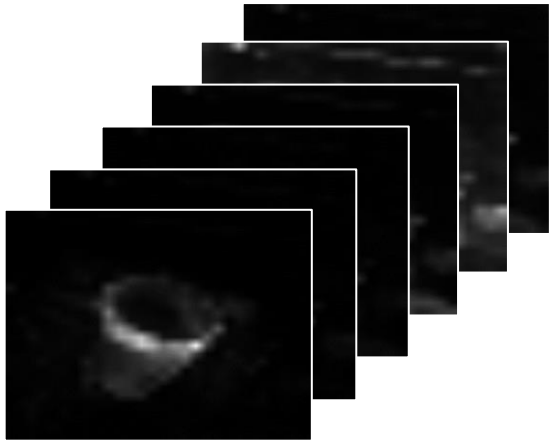
- Dostane:
  - Obrázok 224x224
- Vráti:
  - príznakový vektor 1x384
  - 6 x attention map
- Apache License, dostupný kód i váhy
  dostupný aj v ONNX formáte
  github.com/facebookresearch/dino

- 4GB GPU alebo CPU
- Netrénuje sa
- Inferencia:
  - 0.67s na CUDA s loadovanim modelu
  - 0.02s na CUDA keď už je model naloadovaný
  - 0.13s na CPU
- platforma Pytorch, Ubuntu, odvodený od ViT, vydal Meta AI (Facebook)

DINO (self-DIstillation with NO Labels) v1





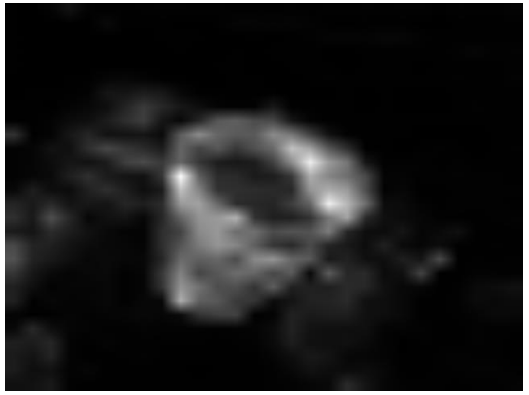
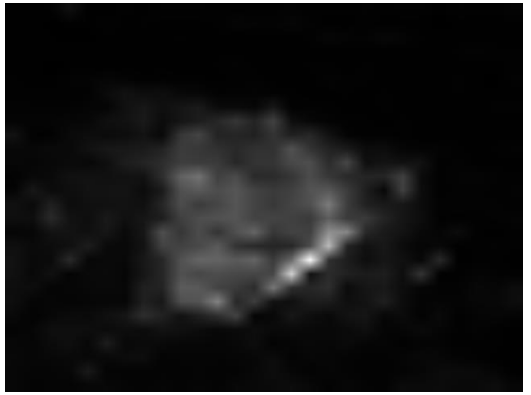


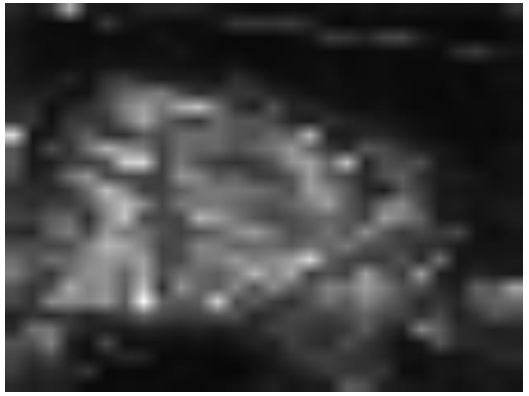
```
[[ 0.37   2.63   0.7    0.66  -0.25  -2.94  -1.57   3.66  -1.14   2.21
  -2.61   6.49   3.55  -3.92  -1.89  12.06  -4.2    1.16   5.12  -2.41
   3.24  -3.82   0.14   5.06  -2.52  10.69  -1.63   1.73  -0.41   1.6
   0.02   0.11   0.33  -0.84  -2.36  -2.96  -4.43   1.32  -1.57   0.03
   2.32   3.31   1.93   1.46  -0.84  14.62  -0.1    0.49  -3.44   1.89
  -0.53   1.04  -2.     1.58  -3.18   0.46  -5.31   1.68   2.17  -4.7
   0.82  -1.25  -0.17  -5.52   1.06   5.82  -2.36  -1.86  -2.2   -1.93
  -5.48  -2.73  -2.02  -0.53  14.55  -4.19   5.7    -2.02   1.1  -10.93
  -0.3   -1.8    0.97   0.63  -4.91  -1.63  -0.21  -5.03  -3.25   7.83
  -4.9   -1.59  -1.32   1.73  -7.65   0.78   3.06  -2.85  -0.43   4.66
   5.16   2.61   5.53   0.82   0.05   3.62  -1.28   0.7    1.87  -1.19
  -8.28   2.16   0.5   -1.17   1.74   2.08  -4.38   6.68   5.02   6.27
  -3.14  12.75 -16.36  -1.21   7.25  -1.63   1.71  -5.21   6.9    1.98
  -1.75   2.8   -3.03   1.3    7.8    3.06  -4.18  -4.35 -12.24  -0.08
  -3.5    2.51  -0.19  -3.81  -4.18   7.67  -2.84   1.41   0.87  -3.27
   0.16  -0.09   1.73  -2.23  -9.82  -2.58  -3.4   -4.08   0.56  -2.48
  -5.39   4.59   0.72   3.32   3.29  -3.6   -0.13   4.65  -5.15  -5.24
  -2.32   5.93   0.89   2.02  -3.25  -1.98  -0.64   4.24   8.09  -5.61
  -3.6   -0.18  -5.54   0.6   -3.88  -5.02   2.02  -1.16   1.77   2.58
  -1.25   0.32  -4.24   3.61  -0.5   -2.89   1.52   7.71  -2.9   10.41
  -3.12  -5.3    4.03   2.    -5.6   -2.29   7.02   3.53   2.36  -2.59
   1.41   5.     2.18  -2.36   3.39   5.55   4.47   1.59   4.22   0.68
   1.92  -0.12  -3.52  -2.86   1.18  -1.92   9.13  -1.04   0.71   3.39
  -5.35  -1.52  -3.46   4.2   -0.22  -0.17   5.58   1.04  -5.02   0.95
  -0.57   4.32  -2.39   2.71  -1.65  -1.62   3.19  -1.44   0.61   2.51
  -2.11   2.93  -0.2  -13.94  -3.31  -5.4    6.45  -3.6    4.73  -2.23
  -1.02   0.57  -1.45   0.89  -3.3   -0.41   2.56 -15.4   -3.78  -6.35
   1.45   9.59  -3.38   6.18  -6.55   4.05  -2.75   3.43  -6.72  -7.45
   6.67  -7.03  -1.58   6.16  -0.84  -0.22   2.63  -2.92  -6.13  -4.14
   1.31  -2.06   1.31   0.02   1.42   1.36  -2.95   7.2   -10.27   1.29
   1.42   1.56   5.59   4.71  -3.84  -0.     2.94  -4.96  -1.7   -0.57
   6.49   4.24   3.33   4.18   0.52   2.82  -3.45  -6.27   1.52  -3.25
  -3.31   4.92   4.1    2.47  -0.99   9.92   2.36  -7.38   3.18   0.93
  -0.37   3.08  -2.4   -0.33  -2.97   5.68   1.38 -10.75   1.02   4.69
   4.61  -4.56   4.14  -4.58   0.44  -6.04  -6.2   -3.05  -3.61  -1.19
  -0.05   1.59   1.01   1.36  -1.4    4.09   4.56   6.13  -1.64  -0.25
  -1.57  -2.04   0.74  -6.78  -3.18   0.09  -0.53   6.95   5.38   4.57
   1.83  -2.76   0.59  -3.79   4.85  -4.15   1.11  -3.35   0.87  -4.42
  -1.34   4.35   7.02  -1.62]]
```

DINO (self-DIstillation with NO Labels) v1

attention (pozornostné) mapy



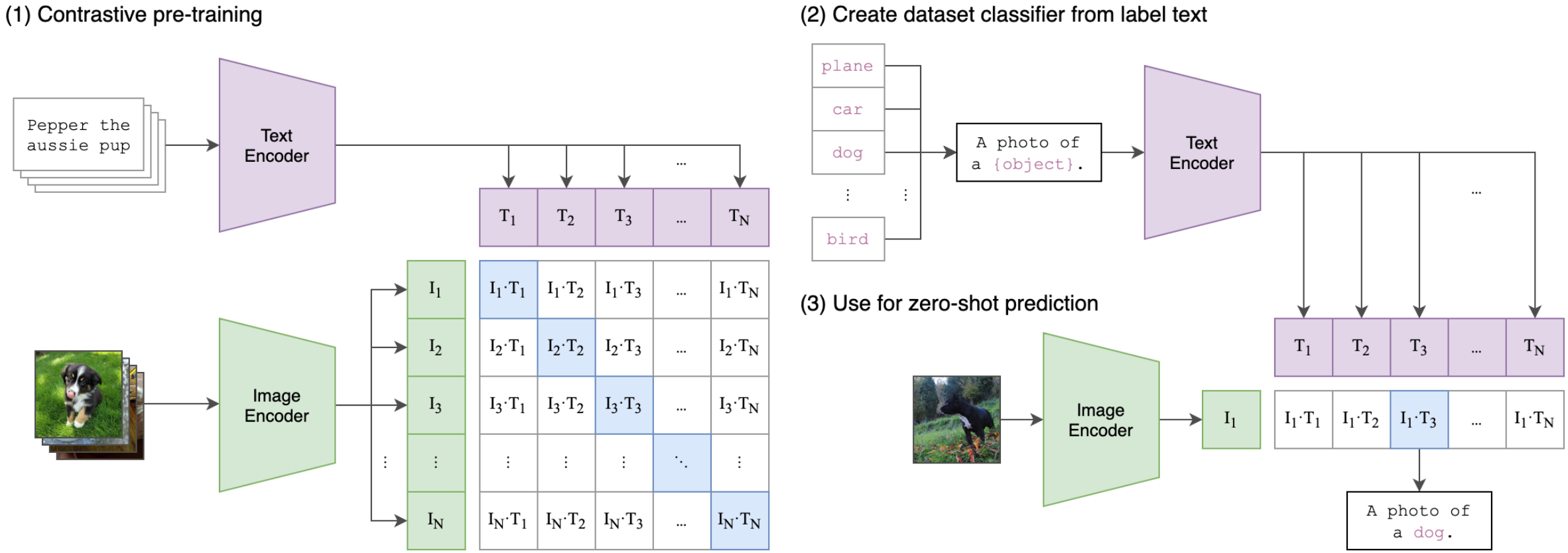






CLIP (Contrastive Language-Image Pre-training) - slov. „záber“ či „klip“

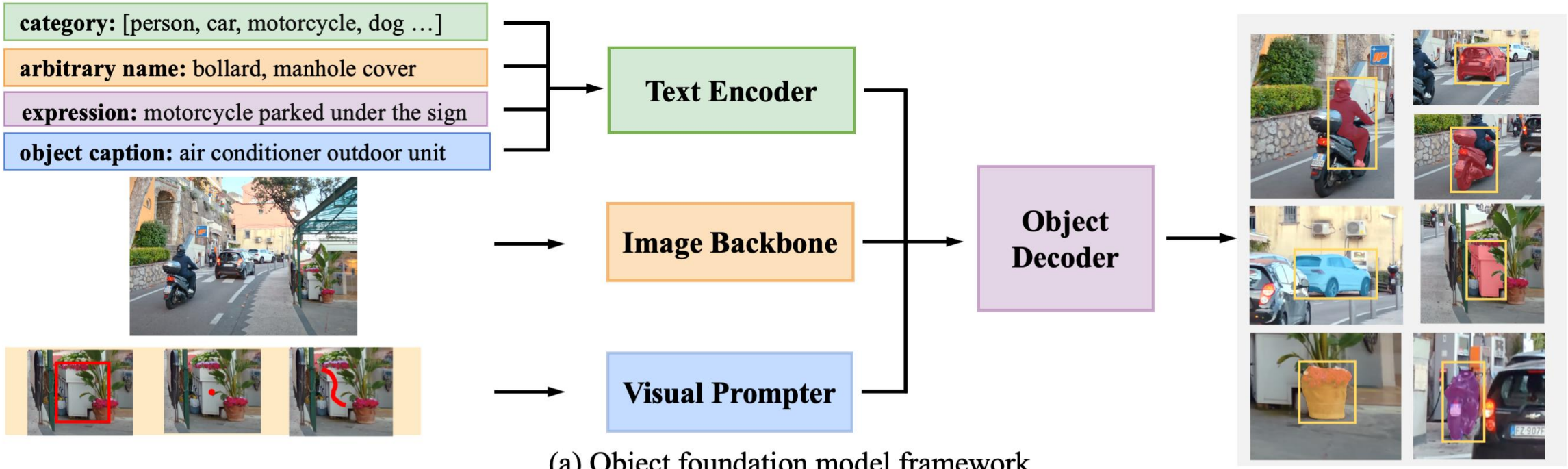
- Dostane:
  - obrázok a texty
- Spočíta pre ne:
  - príznakové vektory
- Odpovie:
  - ktorý z textov najviac zodpovedá obrázku (čo je ten, ktorého vektor má s vektorom obrázka najväčší skalárny súčin
- MIT License, dostupný kód i váhy a to aj v ONNX formáte
  github.com/openai/CLIP
- primárne pre CPU
- Netrénuje sa (zero-shot), vyrovná sa kvalitou ResNet50 natrénovanej na špecifickom datasete
- Inferencia:
  - 0.1s obraz, 0.1s text na CPU
  - 7s loadovanie na CUDA
  - 0.12s obraz, 0.09s text na CUDA
- Platforma: Pytorch / transformers, Ubuntu, odvodený od ViT a Roberta, vydal OpenAI (MicroSoft)

CLIP (Contrastive Language-Image Pre-training)

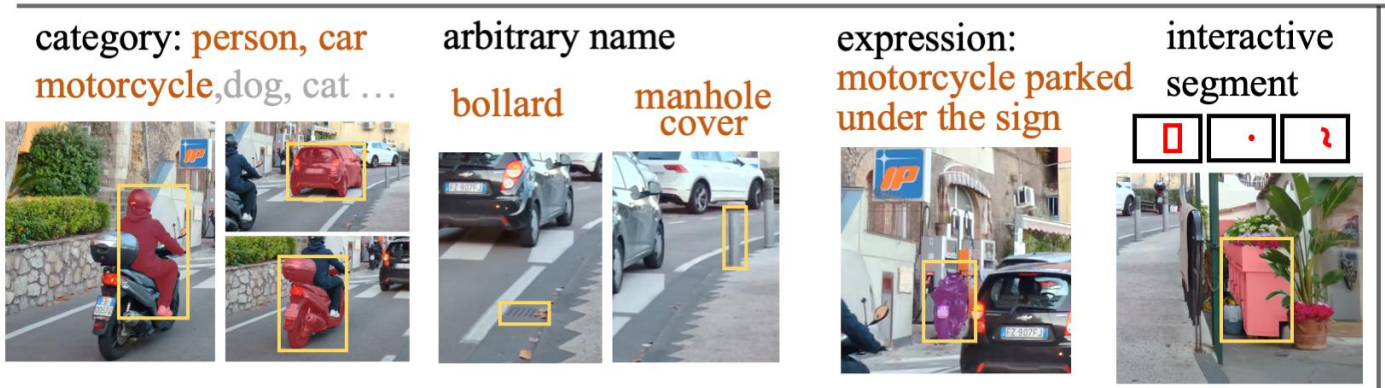


- Model sa skladá z dvoch enkóderov (text a obraz), kódy textov si možno pripraviť vopred a potom sa spočíta kód obrazu a vynásobí sa ich maticou

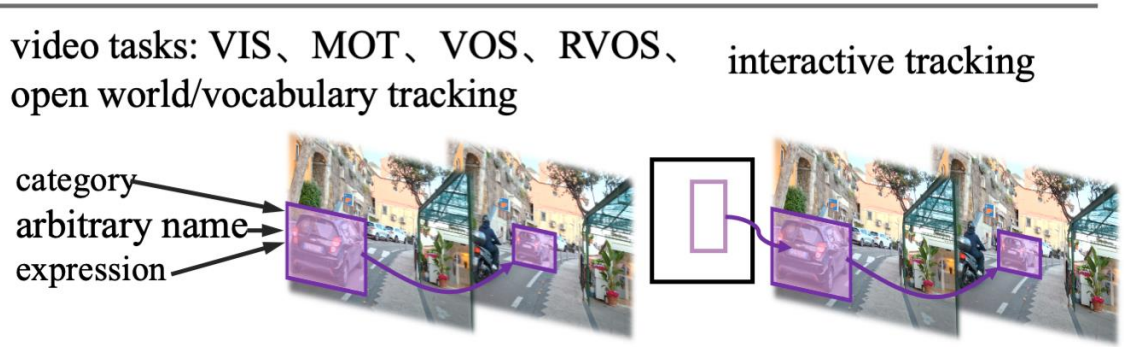
GLEE (General Object Foundation Model for Images and Videos at Scale)



(a) Object foundation model framework



(b) Applied to image tasks



(c) Applied to video tasks

GLEE (General Object Foundation Model for Images and Videos at Scale) - slov. „radosť“

- Dostane obrázok a prompt
- Prompt môže byť:
  - expression
  - point
  - scribble
  - categories
- Odpovie:
  - obdĺžnik
  - maska
  - skóre
- MIT License, dostupný kód i váhy
  github.com/FoundationVision/GLEE
- 8GB GPU alebo CPU
- Netrénuje sa
- Inferencia:
  - 2.96s na CUDA s loadovanim modelu
  - 0.65s na CUDA keď už je model naloadovaný
  - 7.86s na CPU
- platforma Pytorch, Ubuntu, odvodený od Detectron2 a LLaMA, vydal Meta AI (Facebook)
- online:
  https://huggingface.co/spaces/Junfeng5/GLEE_demo

GLEE (General Object Foundation Model for Images and Videos at Scale)

Prompt: The first dog from the right 800 x 4404 8GB, 1.5s