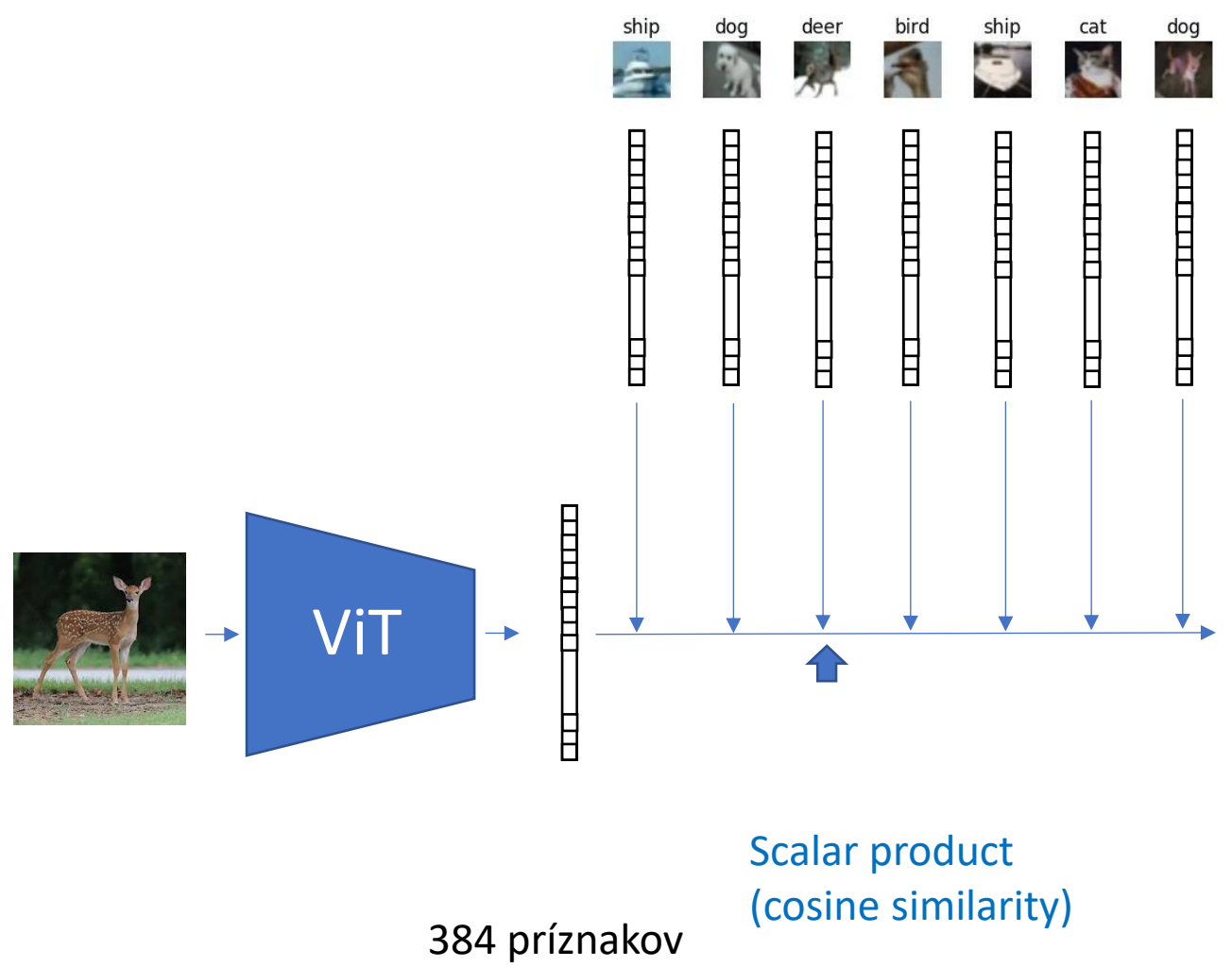


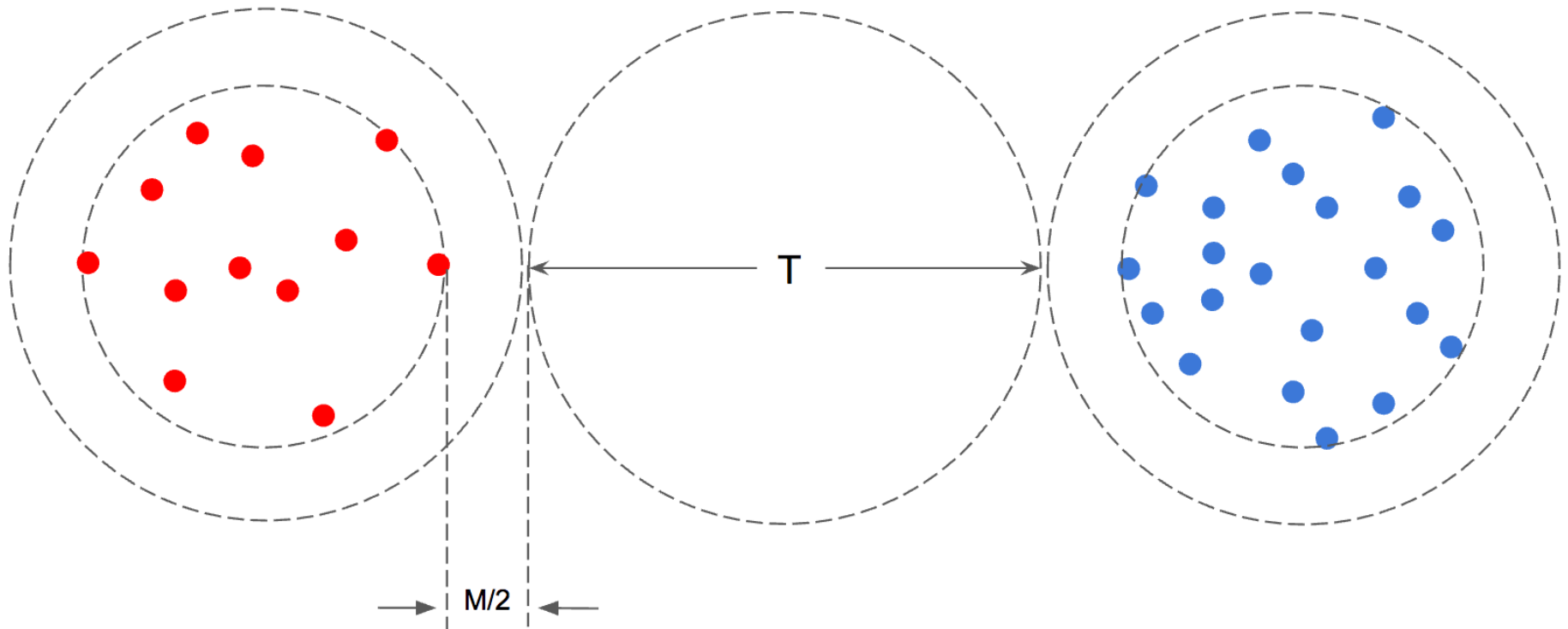
# Enrollment



# Enrollment

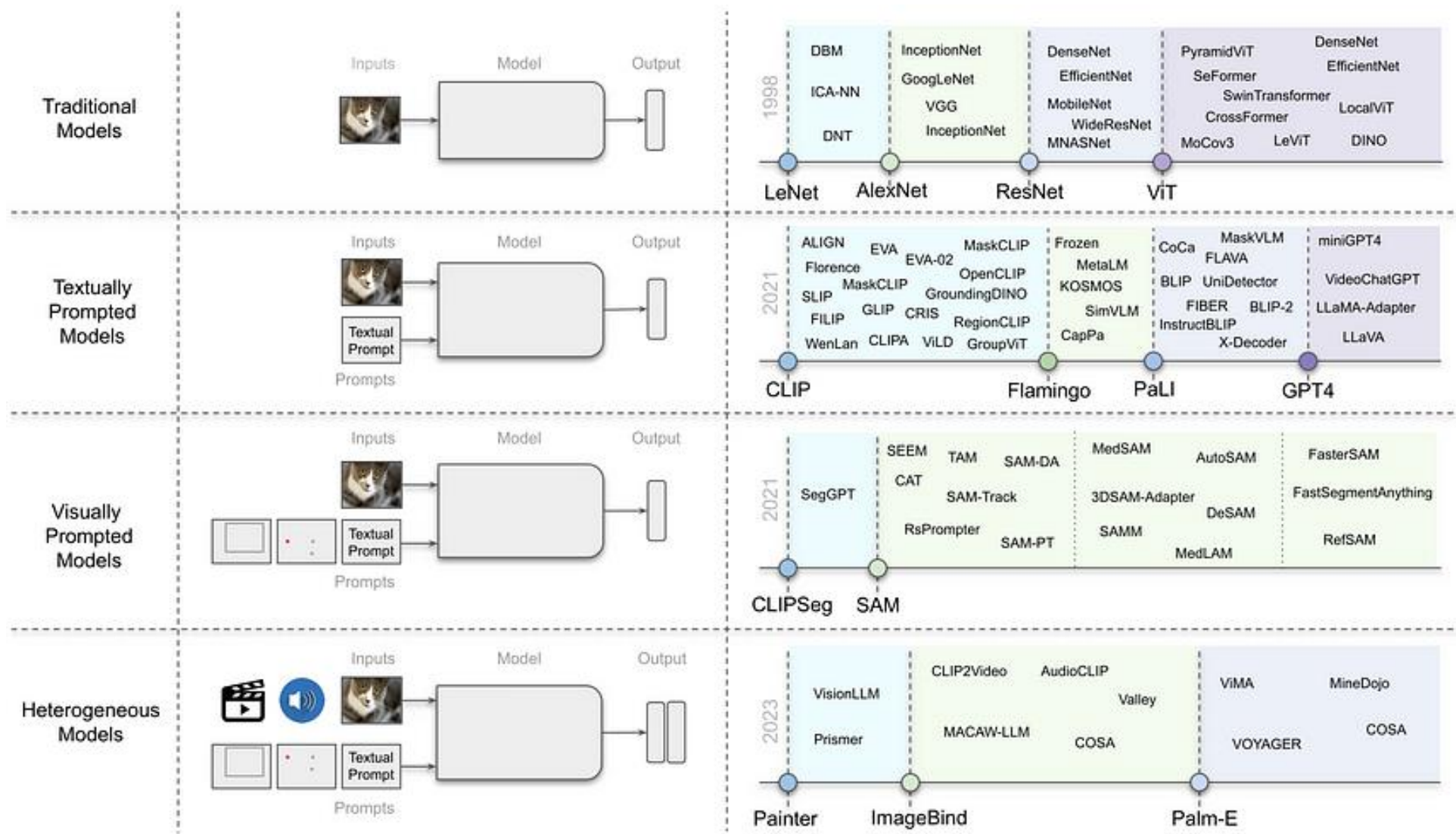
# Contrastive (Metric) Learning

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



- modely, ktoré nie sú určené na tréningovanie 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 natréňované z obrovského množstva dát, z veľkej časti spôsobom self-supervision
- Prístup k strojovému učeniu: data-centric → model-centric

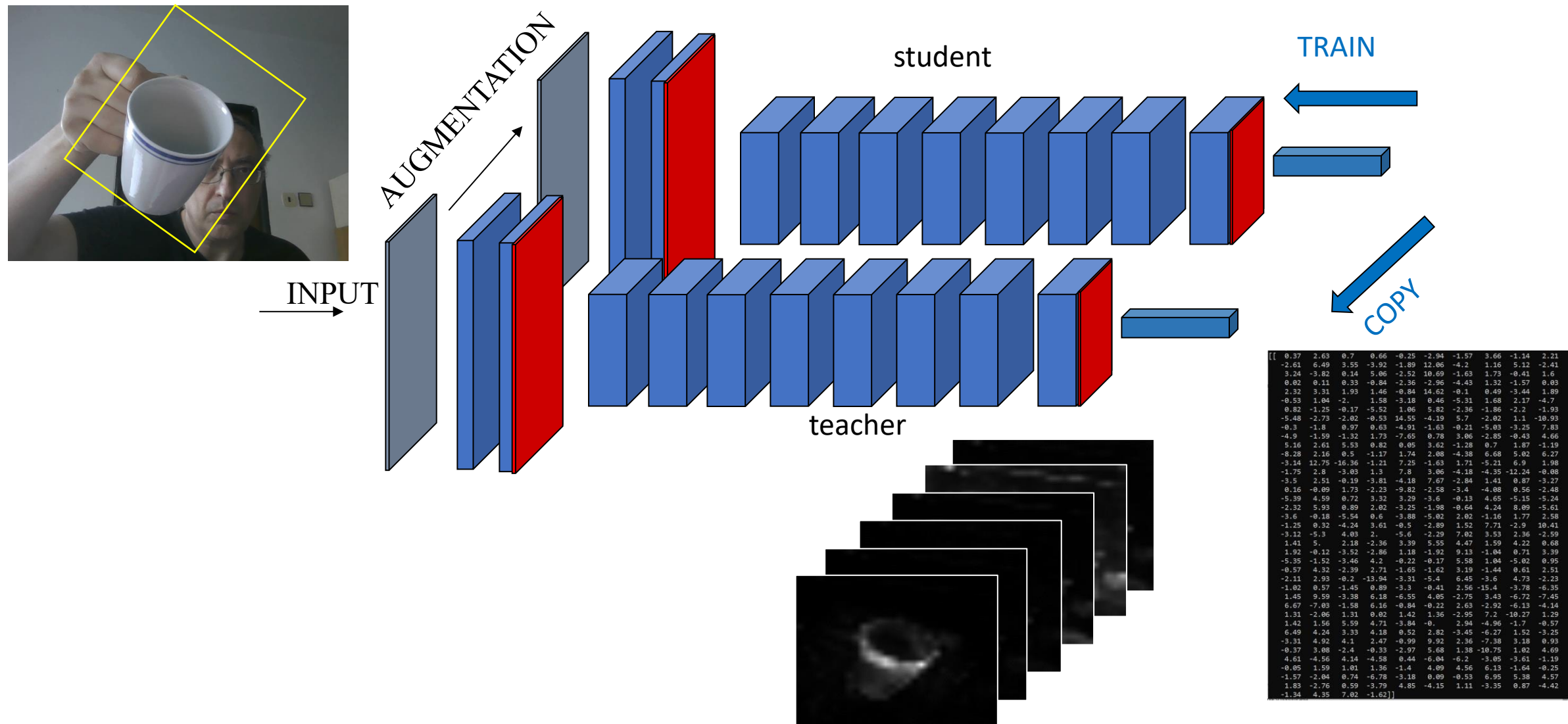
# Foundation modely



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

- 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](https://github.com/facebookresearch/dino)
- 4GB GPU alebo CPU
- Netrénuje sa
- Inferencia:
  - 0.67s na CUDA s loadovaním modelu
  - 0.02s na CUDA keď už je model naložovaný
  - 0.13s na CPU
- platforma Pytorch, Ubuntu, odvodený od ViT, vydal Meta AI (Facebook)

# DINO (self-Distillation with NO Labels) v1



attention (pozornostné) mapy

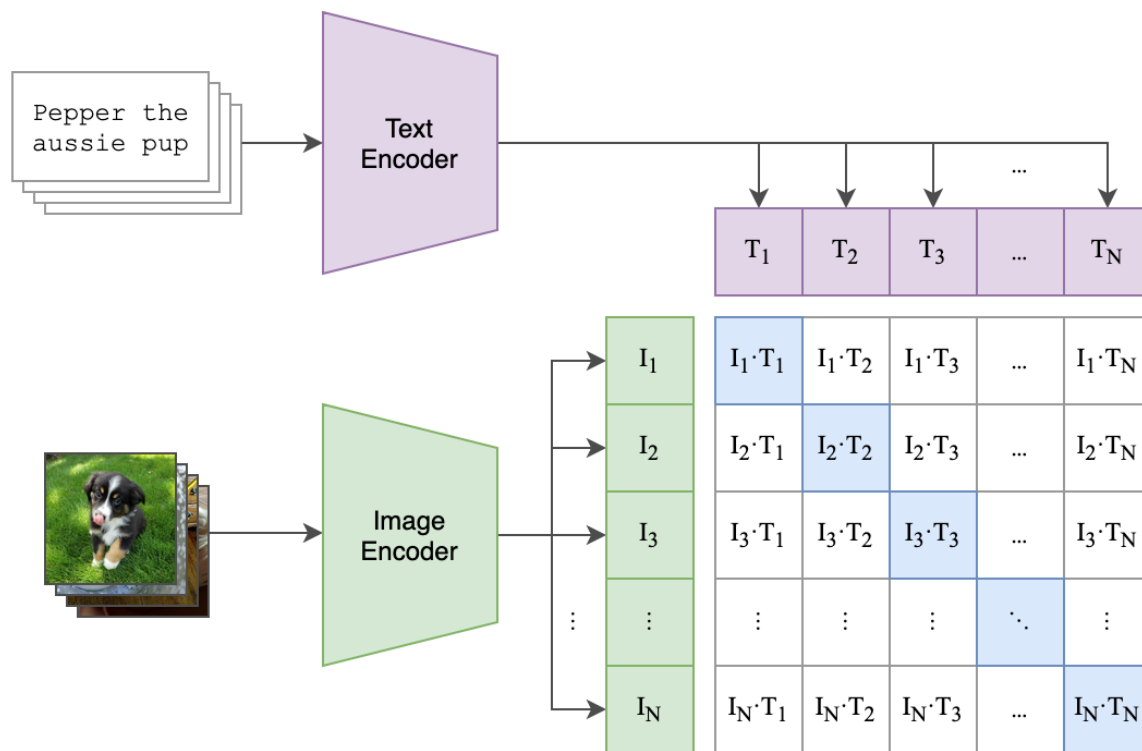




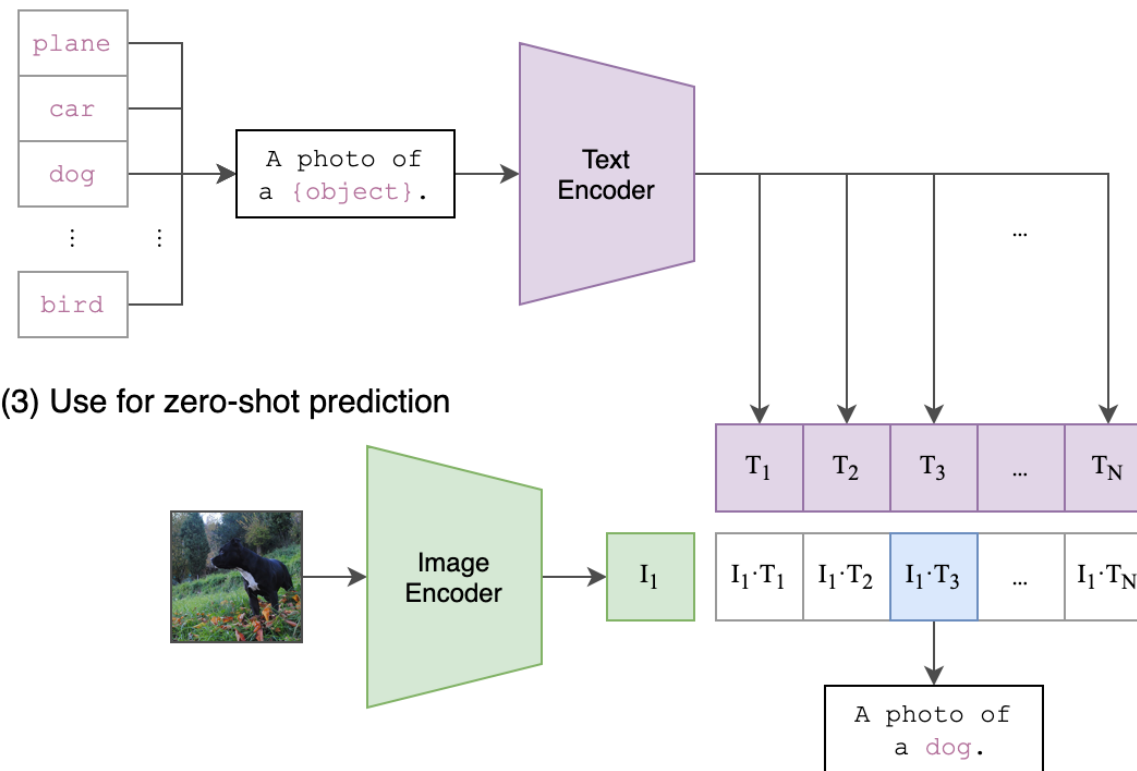
- Dostane:
  - obrázkov a texty
- Spočíta pre ne:
  - príznakové vektory
- Odpovie:
  - ktorý z textov najviac zodpovedá obrázku (čo je ten, ktorého vektor má s vektorom obrázka najväčší skalárny súčin)
- MIT License, dostupný kód i váhy a to aj v ONNX formáte  
[github.com/openai/CLIP](https://github.com/openai/CLIP)
- 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)

(1) Contrastive pre-training

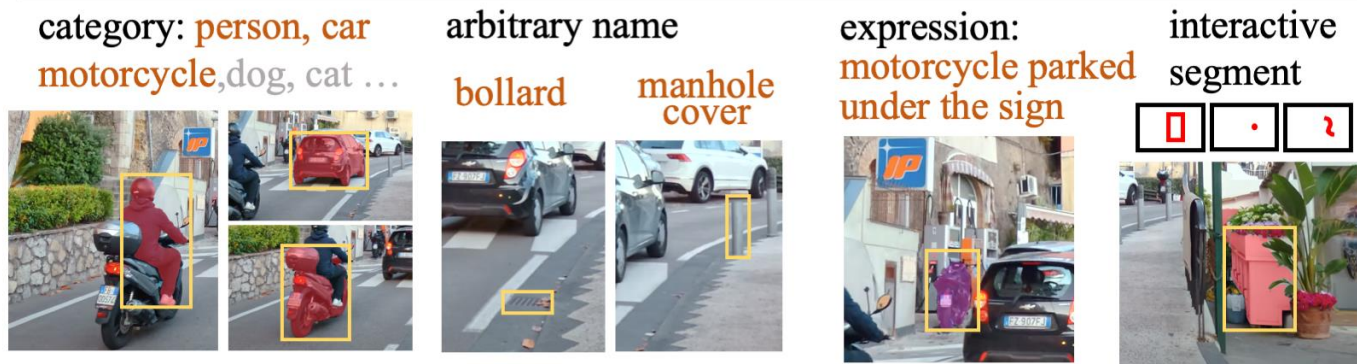
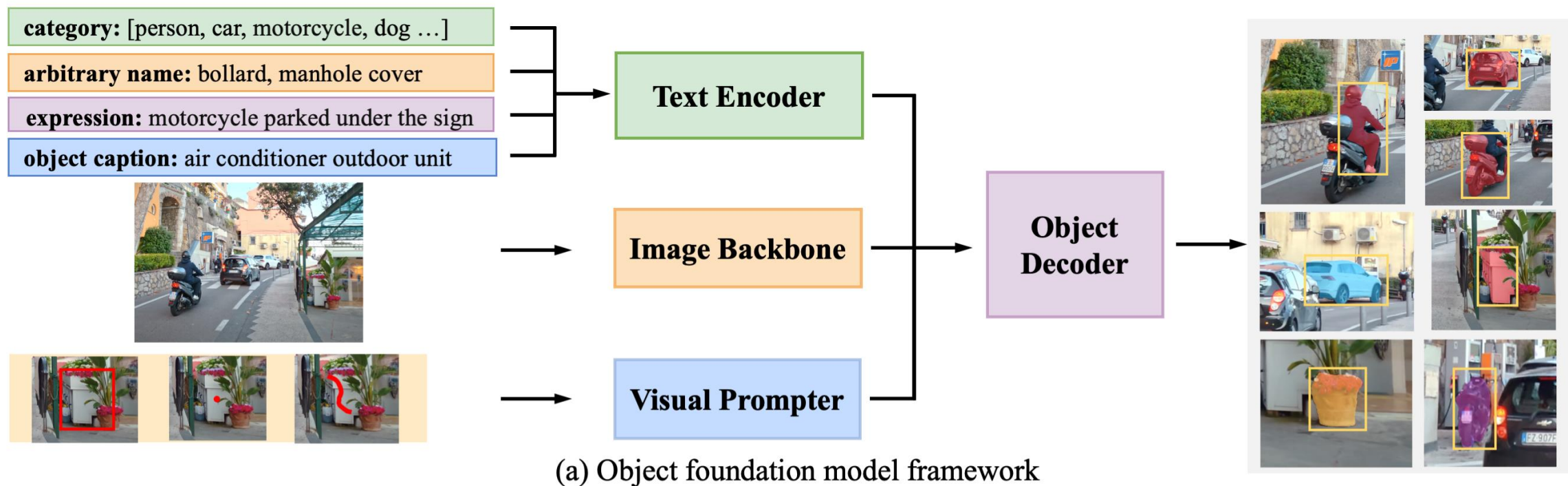


(2) Create dataset classifier from label text



- 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)



(b) Applied to image tasks

video tasks: VIS、MOT、VOS、RVOS、open world/vocabulary tracking      interactive tracking



(c) Applied to video tasks

- 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](https://github.com/FoundationVision/GLEE)
- 8GB GPU alebo CPU
- Netrénuje sa
- Inferencia:
  - 2.96s na CUDA s loadovaním modelu
  - 0.65s na CUDA keď už je model naložovaný
  - 7.86s na CPU
- platforma Pytorch, Ubuntu, odvodený od Detectron2 a LLaMA, vydal Meta AI (Facebook)
- online:  
[https://huggingface.co/spaces/Junfeng5/GLEE\\_demo](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

