

Jak hledat klip CLIPem?

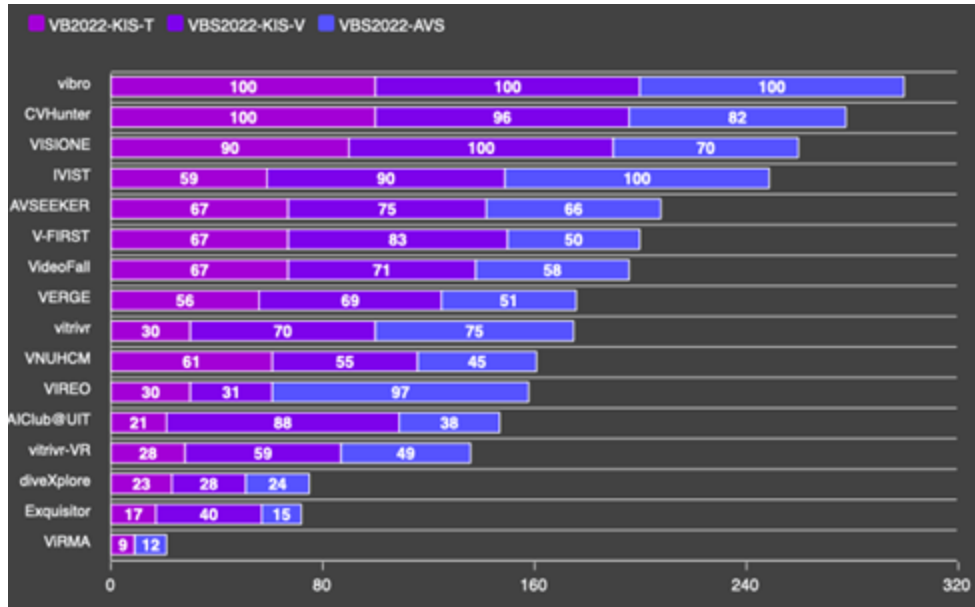
Jakub Lokoč



matfyz

Soutěž od 17:30 do 18:00 (o ceny do 18:30)

- Tři typy úloh pro soutěž:
vizuální, textové a Q/A
- Motivační odměna pro ty,
kteří zůstanou do 18:30
3 mikiny, 7 lahví (na vodu !!)





Dotaz: “Kdo rád pracuje horizontálně?”



10 minut? Hodin? Dnů? Měsíců? Roků?

snadné

složité?



1	2	2	1	3	5	2	...
1	2	2	2	3	4	1	...
0	1	1	1	2	3	2	...
1	1	1	2	3	4	3	...
1	1	1	1	2	3	4	...
1	1	1	2	3	4	4	...
1	2	2	1	2	3	4	...
...



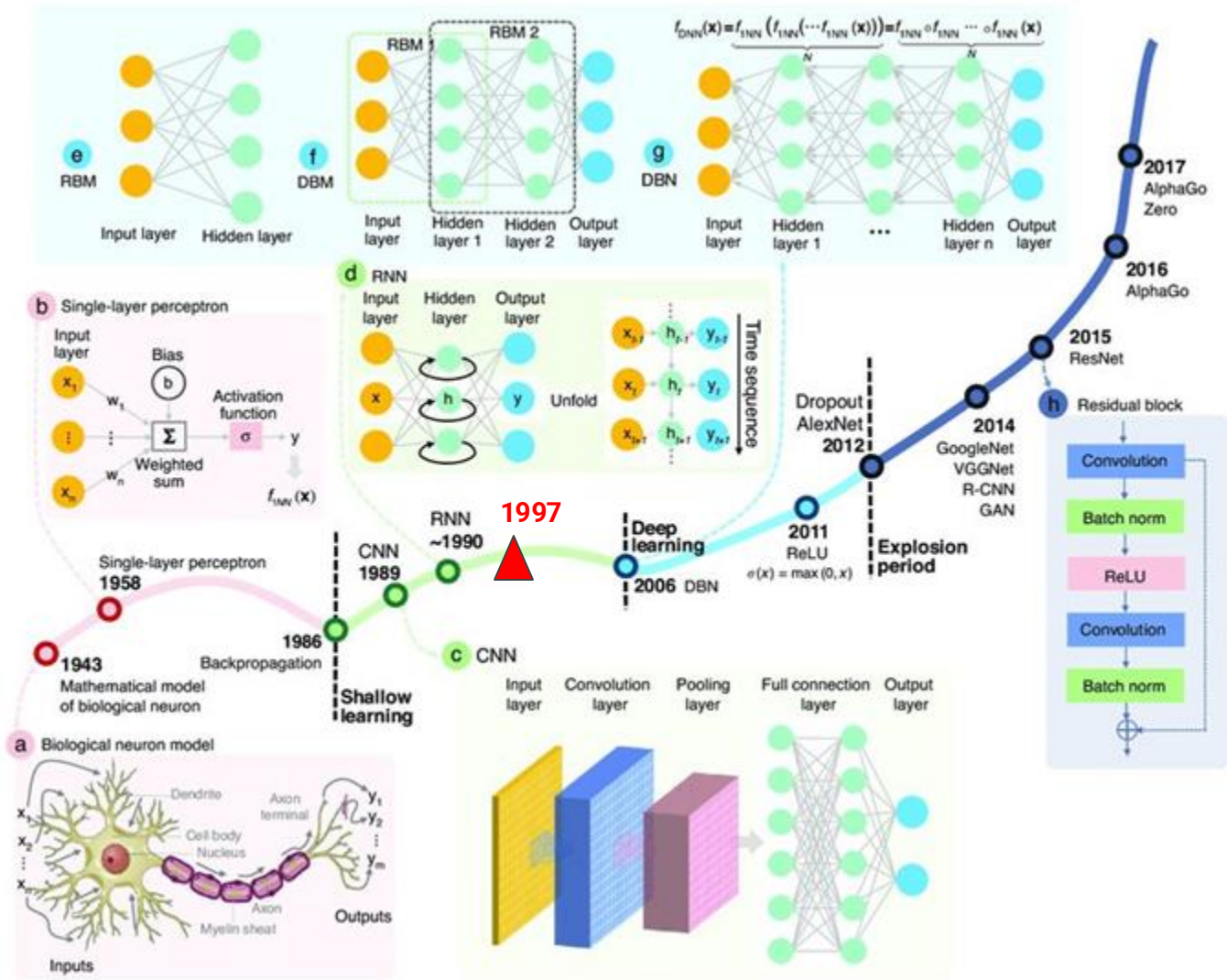


1	2	2	1	3	5	2	...
1	2	2	2	3	4	1	...
0	1	1	1	2	3	2	...
1	1	1	2	3	4	3	...
1	1	1	1	2	3	4	...
1	1	1	2	3	4	4	...
1	2	2	1	2	3	4	...
...

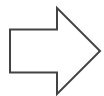


Kůň?

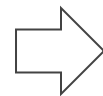
Zuo, Chao & Qian, Jiaming & Feng, Shijie & Yin, Wei & Li, Yixuan & Fan, Pengfei & Han, Jing & Qian, Kemao & Chen, Qian. (2022). Deep learning in optical metrology: a review. Light: Science & Applications. 11. 39. 10.1038/s41377-022-00714-x.



Joint-embedding přístup

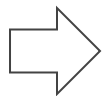


Natrénovaná
síť pro obraz

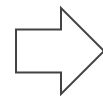


Společný prostor

$x = [1.2, 2.8, \dots, 0.2]$



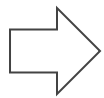
Natrénovaná
síť pro text



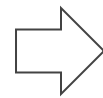
$y = [1.1, 2.6, \dots, 0.4]$

“Skupina koní se
pase na trávě, pak
si jeden kůň lehne
a pak sedne”

Joint-embedding přístup



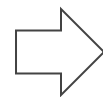
Natrénovaná
síť pro obraz



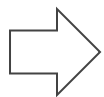
Společný prostor

$x = [1.2, 2.8, \dots, 0.2]$

1000 dimenzí!



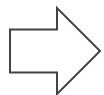
“Skupina koní se
pase na trávě, pak
si jeden kůň lehne
a pak sedne”



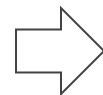
Natrénovaná
síť pro text

$y = [1.1, 2.6, \dots, 0.4]$

Joint-embedding přístup



$f_{\text{obraz}}(o, W)$

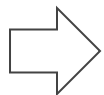


Společný prostor

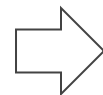
$x = [1.2, 2.8, \dots, 0.2]$

1000 dimenzí!

“Skupina koní se
pase na trávě, pak
si jeden kůň lehne
a pak sedne”



$f_{\text{text}}(t, W')$

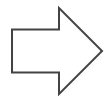


$y = [1.1, 2.6, \dots, 0.4]$

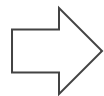
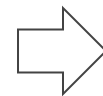
Joint-embedding přístup



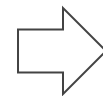
“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”



$f_{\text{obraz}}(o, W)$



$f_{\text{text}}(t, W')$



Společný prostor

$x = [1.2, 2.8, \dots, 0.2]$

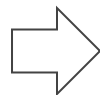
1000 dimenzí!

$y = [1.1, 2.6, \dots, 0.4]$

Grok 3 má 2.7 bilionu parametrů...

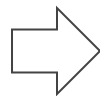
Joint-embedding přístup

$f_{\text{obraz}}(o, W)$



$x = [1.2, 2.8, \dots, 0.2]$

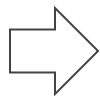
$f_{\text{text}}(t, W')$



$y = [1.1, 2.6, \dots, 0.4]$

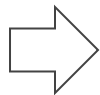
Příklad pro 2D...

$f_{\text{obraz}}(o, W)$

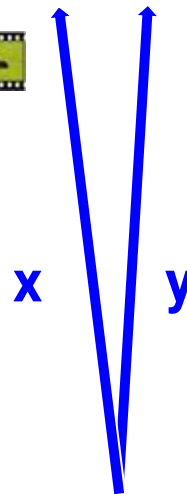


$$x = [-0.2, 2.8]$$

$f_{\text{text}}(t, W')$



$$y = [0.1, 2.6]$$

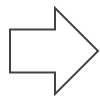


“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

Příklad pro 2D...

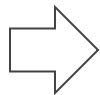
$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

$f_{\text{obraz}}(o, W)$

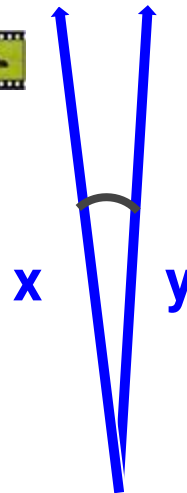


$x = [-0.2, 2.8]$

$f_{\text{text}}(t, W')$



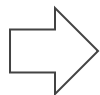
$y = [0.1, 2.6]$



“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

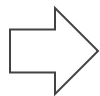
Příklad pro 2D...

$f_{\text{obraz}}(o, W)$



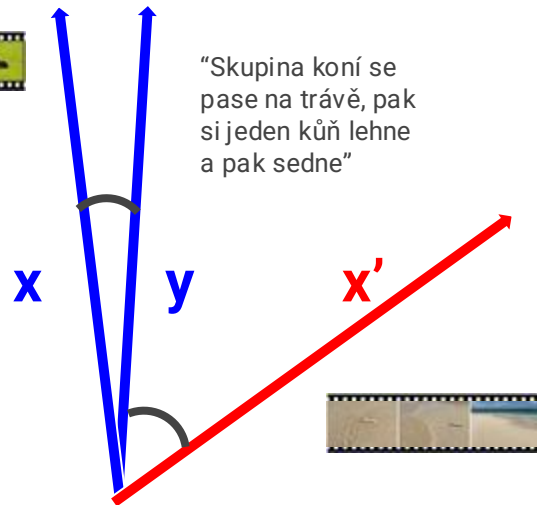
$x = [-0.2, 2.8]$
 $x' = [2.1, 1.1]$

$f_{\text{text}}(t, W')$



$y = [0.1, 2.6]$

$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

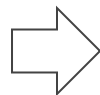


“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

Jakou má roli strojové učení?

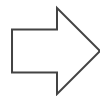
$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

$f_{\text{obraz}}(o, \mathbf{W})$

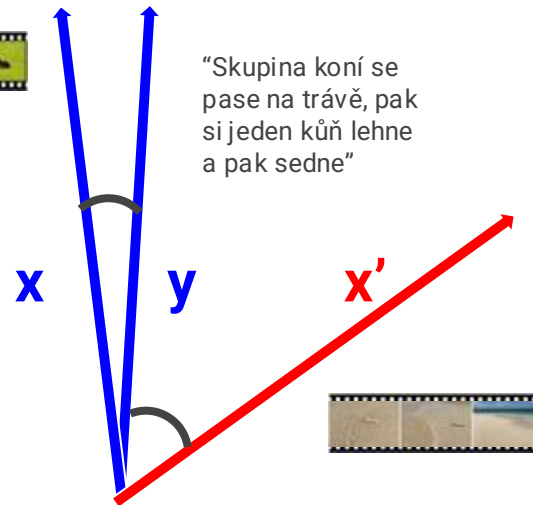


$x = [-0.2, 2.8]$
 $x' = [2.1, 1.1]$

$f_{\text{text}}(t, \mathbf{W}')$



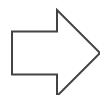
$y = [0.1, 2.6]$



Jakou má roli strojové učení?

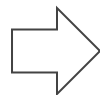
$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

$f_{\text{obraz}}(o, \mathbf{W})$

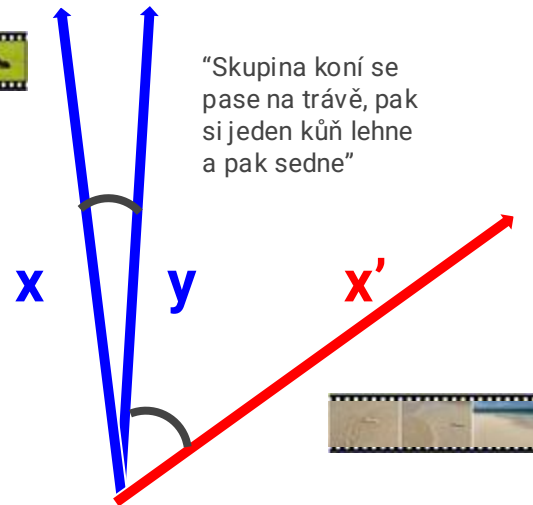


$x = [-0.2, 2.8]$
 $x' = [2.1, 1.1]$

$f_{\text{text}}(t, \mathbf{W}')$



$y = [0.1, 2.6]$

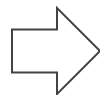


$$\text{Max}(\cos(\mathbf{x}', \mathbf{y}) - \cos(\mathbf{x}, \mathbf{y}) + a, 0)$$

Jakou má roli strojové učení?

$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

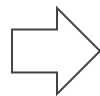
$f_{\text{obraz}}(o, \mathbf{W})$



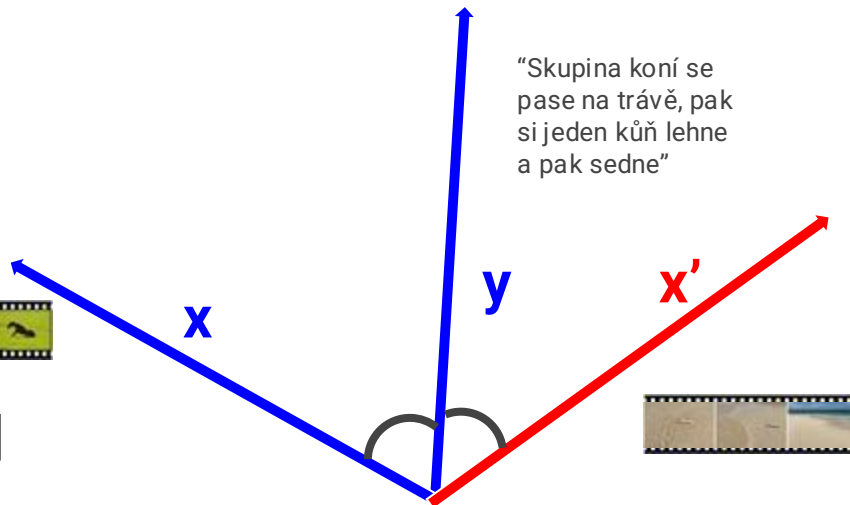
$x = [-1.2, 1.3]$
 $x' = [2.1, 1.1]$



$f_{\text{text}}(t, \mathbf{W}')$



$y = [0.1, 2.6]$

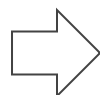


$$\text{Max}(\cos(\mathbf{x}', \mathbf{y}) - \cos(\mathbf{x}, \mathbf{y}) + a, 0)$$

Jakou má roli strojové učení?

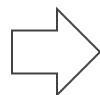
$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

$f_{\text{obraz}}(o, \mathbf{W})$

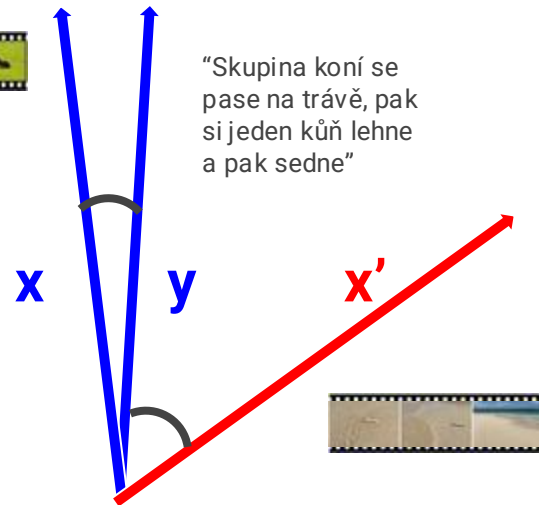


$x = [-0.2, 2.8]$
 $x' = [2.1, 1.1]$

$f_{\text{text}}(t, \mathbf{W}')$



$y = [0.1, 2.6]$



Find \mathbf{W}, \mathbf{W}' to minimize: $\text{Max}(\cos(\mathbf{x}', \mathbf{y}) - \cos(\mathbf{x}, \mathbf{y}) + a, 0) \forall \mathbf{x}, \mathbf{y}, \mathbf{x}'$

Jak využít pro hledání?

“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”



Jak využít pro hledání?

“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

$f_{\text{obraz}}(o, W)$



$x' = [1.2, 2.8, \dots, 0.2]$



$x = [1.2, 2.8, \dots, 0.2]$



$x'' = [1.2, 2.8, \dots, 0.2]$

Jak využít pro hledání?

“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

$f_{\text{text}}(t, W')$

$f_{\text{obraz}}(o, W)$

$y = [1.1, 2.6, \dots, 0.4]$



$x' = [1.2, 2.8, \dots, 0.2]$



$x = [1.2, 2.8, \dots, 0.2]$



$x'' = [1.2, 2.8, \dots, 0.2]$

Jak využít pro hledání?

“Skupina koní se pase na trávě, pak si jeden kůň lehne a pak sedne”

$f_{\text{text}}(t, W')$

$f_{\text{obraz}}(o, W)$



$$x' = [1.2, 2.8, \dots, 0.2]$$



$$x = [1.2, 2.8, \dots, 0.2]$$



$$x'' = [1.2, 2.8, \dots, 0.2]$$

$$y = [1.1, 2.6, \dots, 0.4]$$

$$\cos(x', y) = 0.1$$

$$\cos(x, y) = \mathbf{0.8}$$

$$\cos(x'', y) = 0.2$$

```
import torch
from PIL import Image
import open_clip

model, _, preprocess = open_clip.create_model_and_transforms('ViT-B-32', pretrained='laion2b_s34b_b79k')
model.eval()
tokenizer = open_clip.get_tokenizer('ViT-B-32')

image = preprocess(Image.open("docs/CLIP.png")).unsqueeze(0)
text = tokenizer(["a diagram", "a dog", "a cat"])

with torch.no_grad(), torch.cuda.amp.autocast():
    image_features = model.encode_image(image)
    text_features = model.encode_text(text)
    image_features /= image_features.norm(dim=-1, keepdim=True)
    text_features /= text_features.norm(dim=-1, keepdim=True)

text_probs = (100.0 * image_features @ text_features.T).softmax(dim=-1)
```

Výsledky na základě $\cos(x, y)$



Select User **user01** | Eval. ID

Task ID | Select Data... **MVK**

RESET

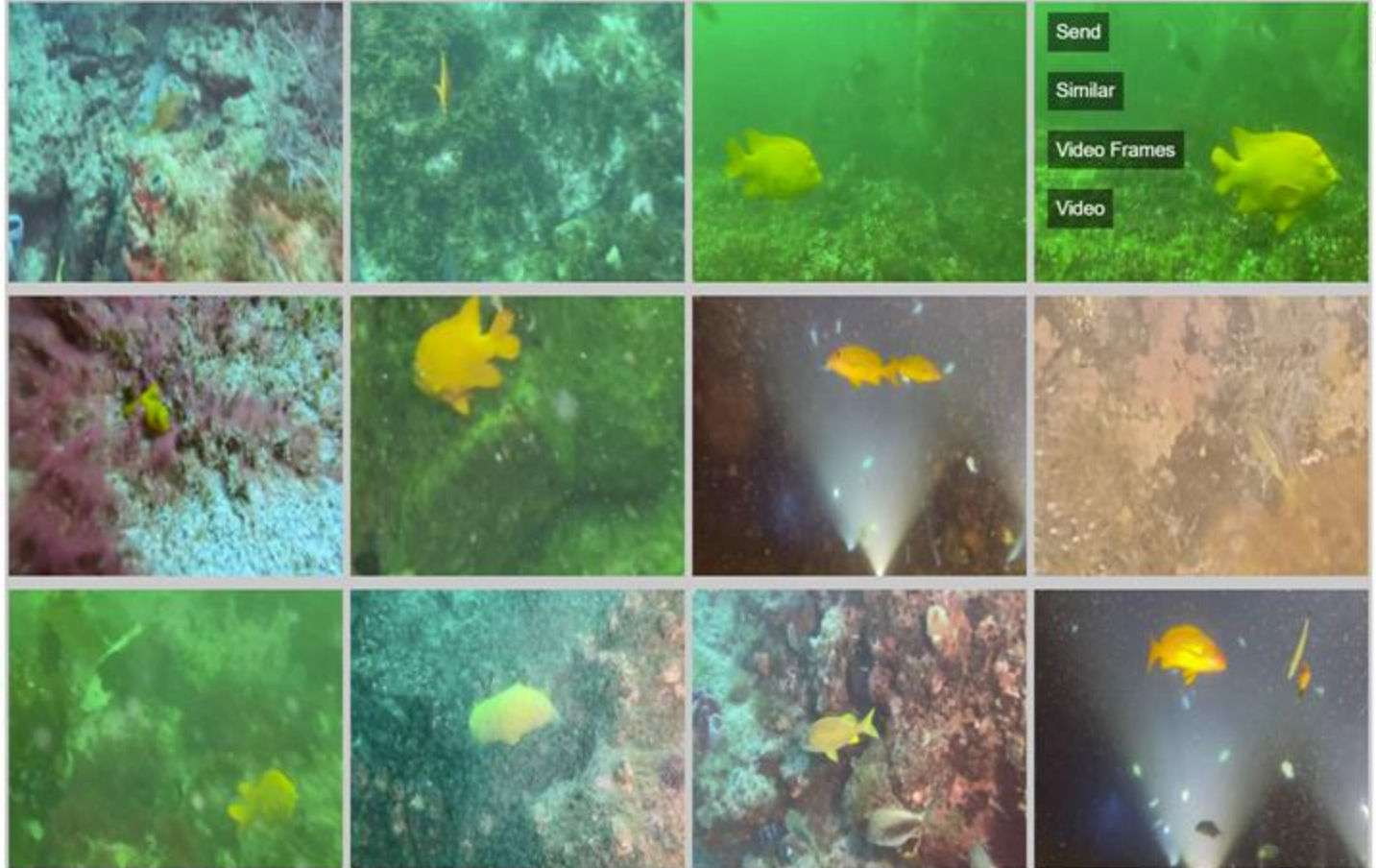
← **PREV** | **NEXT** →

Query Scene 1
yellow fish ← **Dotaz** ⓘ

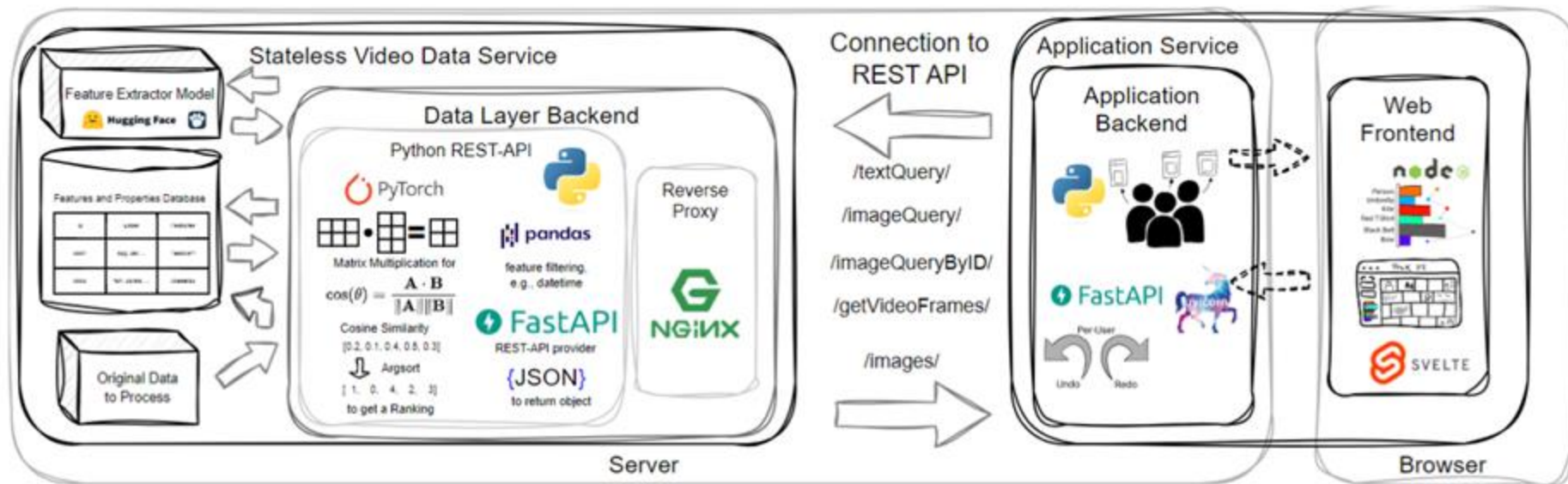
Query Sc... ⓘ

SUBMIT TEXT QUERY

BAYES UPDATE



Proč studovat souběžně AI a vývoj SW?



Proč studovat souběžně AI a vývoj SW?

