

# Active Image Indexing

ICLR 2023, Kigali

Pierre Fernandez, Matthijs Douze, Hervé Jégou, Teddy Furon

pfz@meta.com

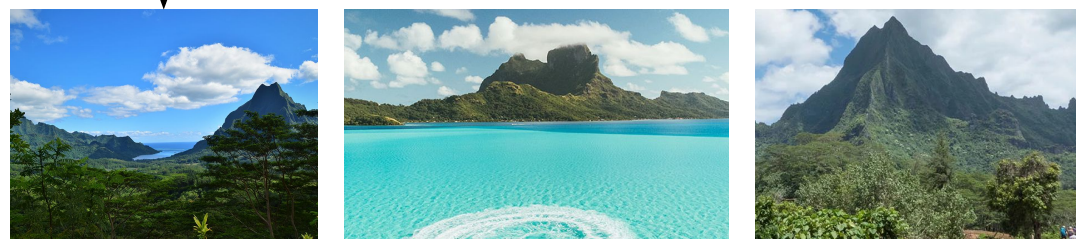
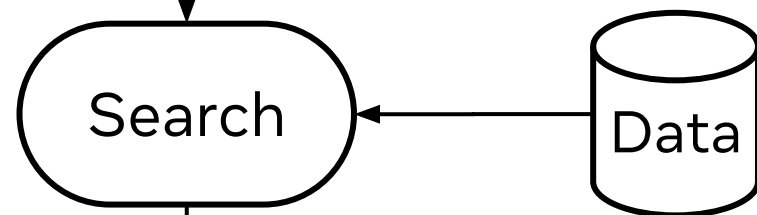


# Introduction

# What is similarity search & copy detection?

- Goal:

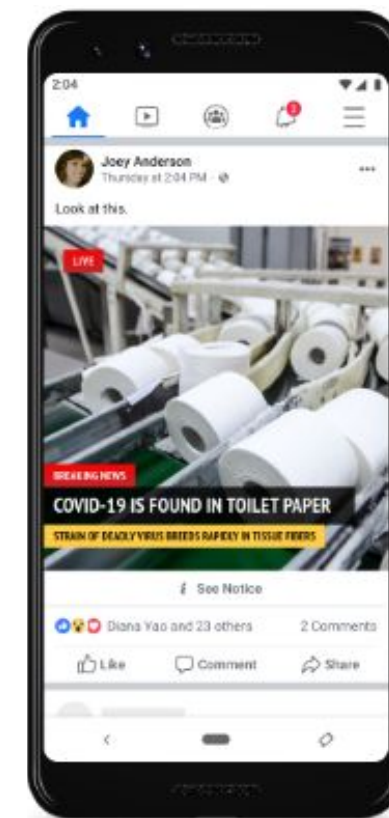
Query



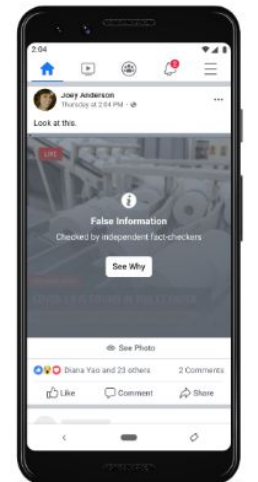
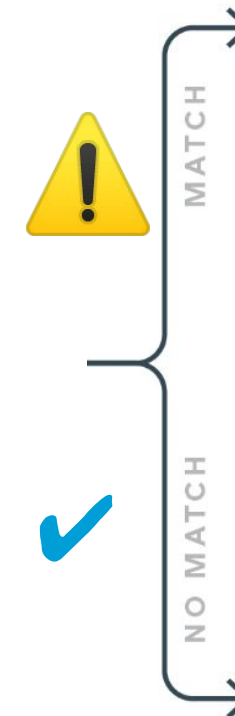
Similar images (& distance)

- Applications

- Intellectual property protection
- De-duplication in search results
- Moderation (Social networks, etc.)



Top candidates that look similar to query



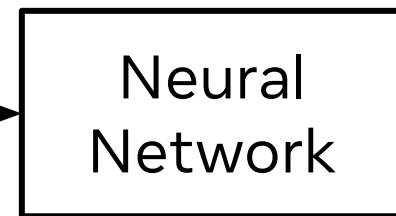
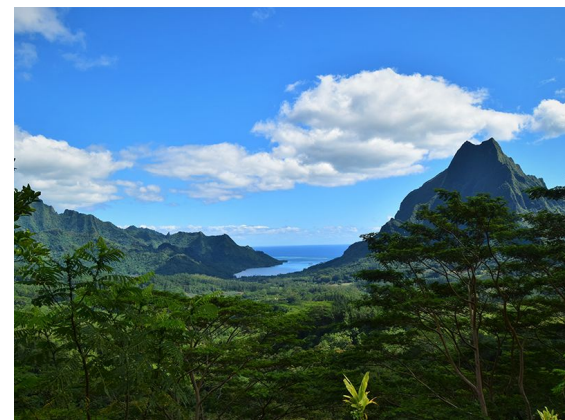
Action taken



No action taken

# How? Representation Learning & Indexing

## (1) Feature extraction



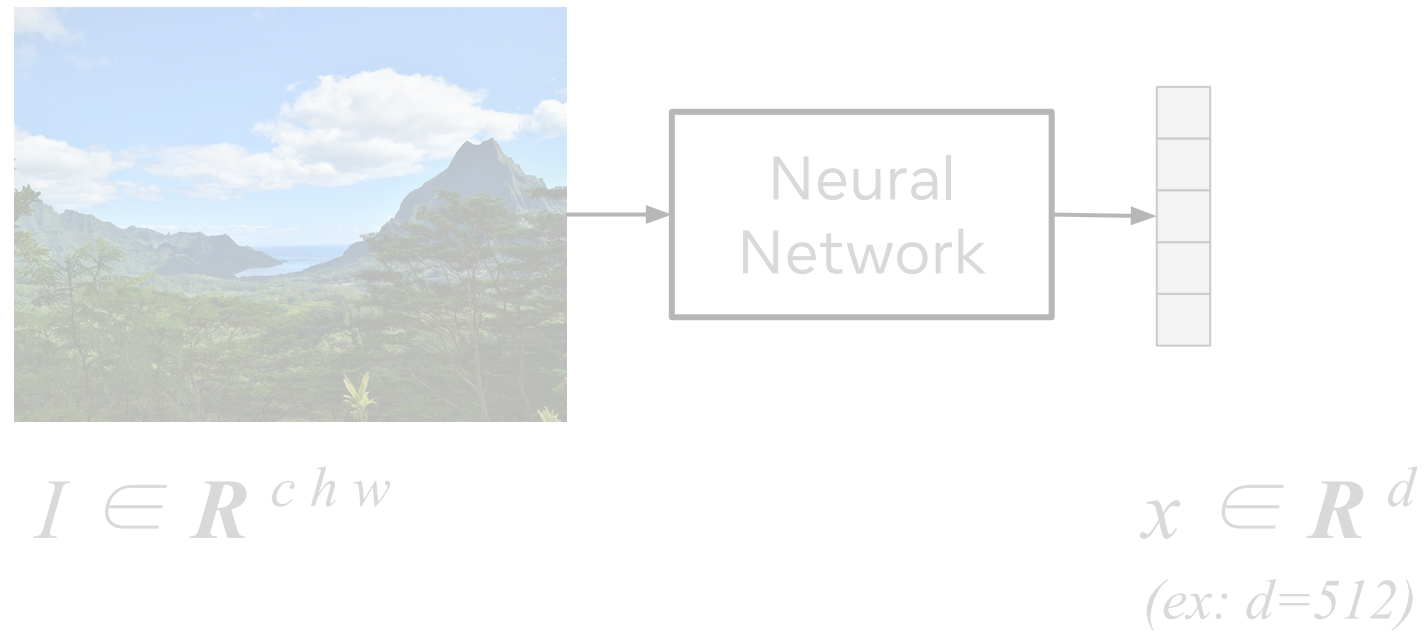
$$I \in \mathbf{R}^{c h w}$$

$$x \in \mathbf{R}^d$$

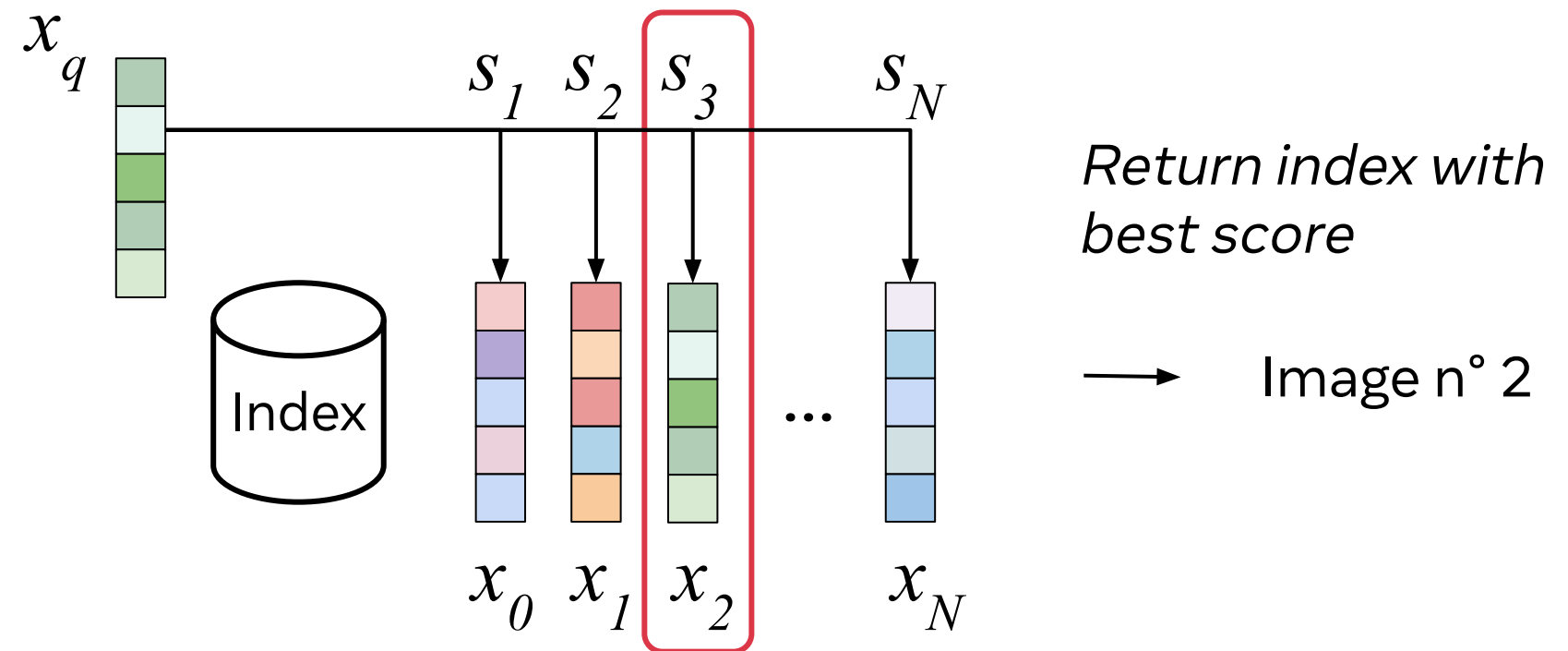
(ex:  $d=512$ )

# How? Representation Learning & Indexing

## (1) Feature extraction



## (2) Similarity Search



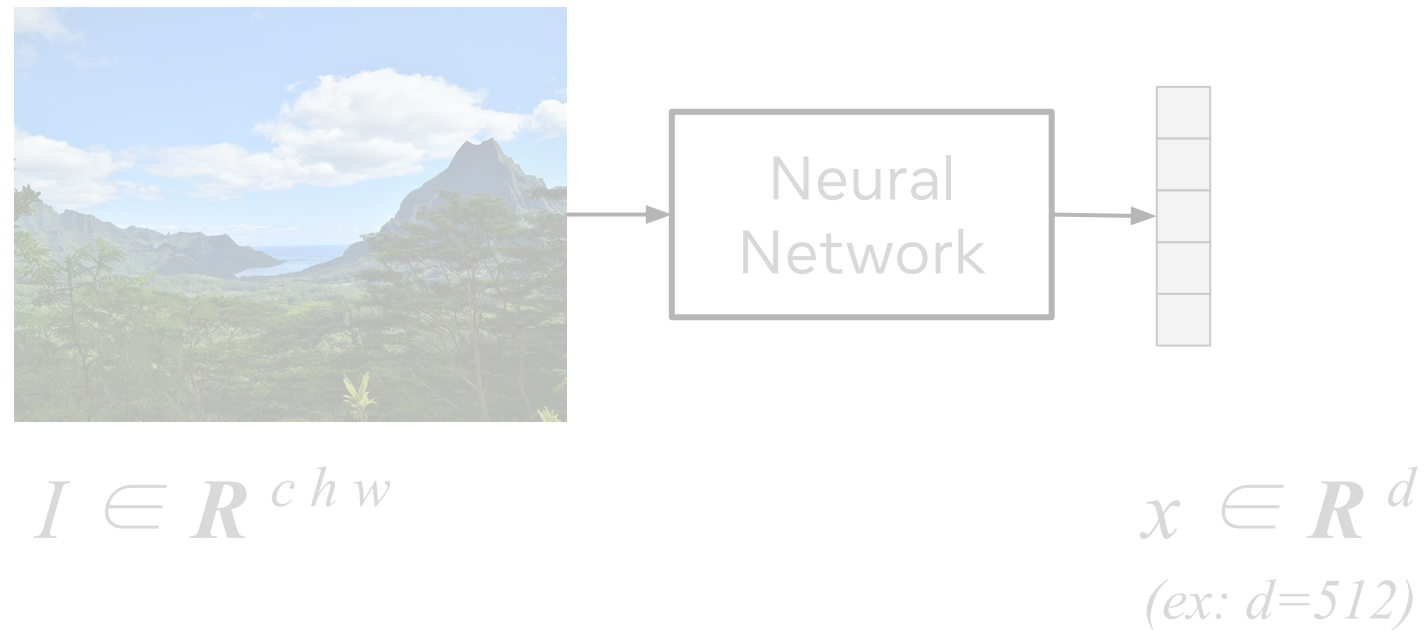
Brute-force search in all embeddings (Flat Index):

→ slow, high memory cost

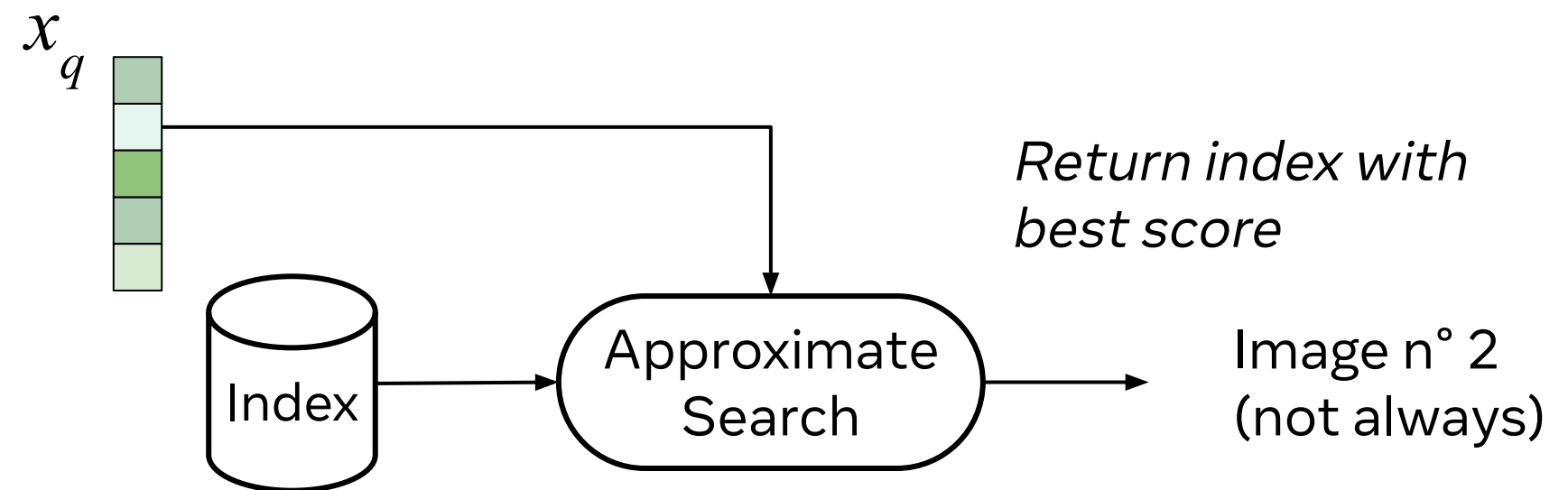
Ex: 1M database images and  $x \in \mathbb{R}^{512}$   
Search time: ¼ s / image  
Memory cost: 2Gb

# How? Representation Learning & Indexing

## (1) Feature extraction



## (2) Approximate Similarity Search



Index (with approximation):

→ less accurate - fast, low memory cost

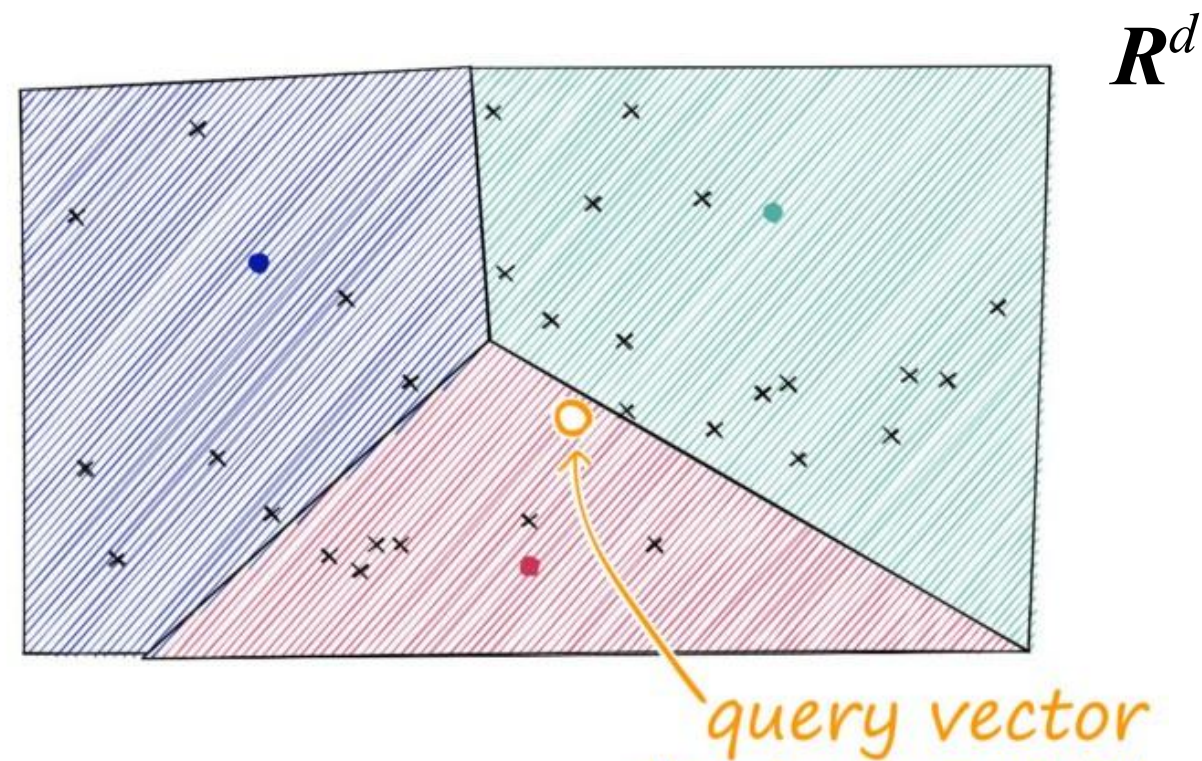
Ex: 1M database images and  $x \in \mathbb{R}^{512}$   
Search time: < 1ms / image  
Memory cost: 8Mb



# Approximations

Ex: space partitioning

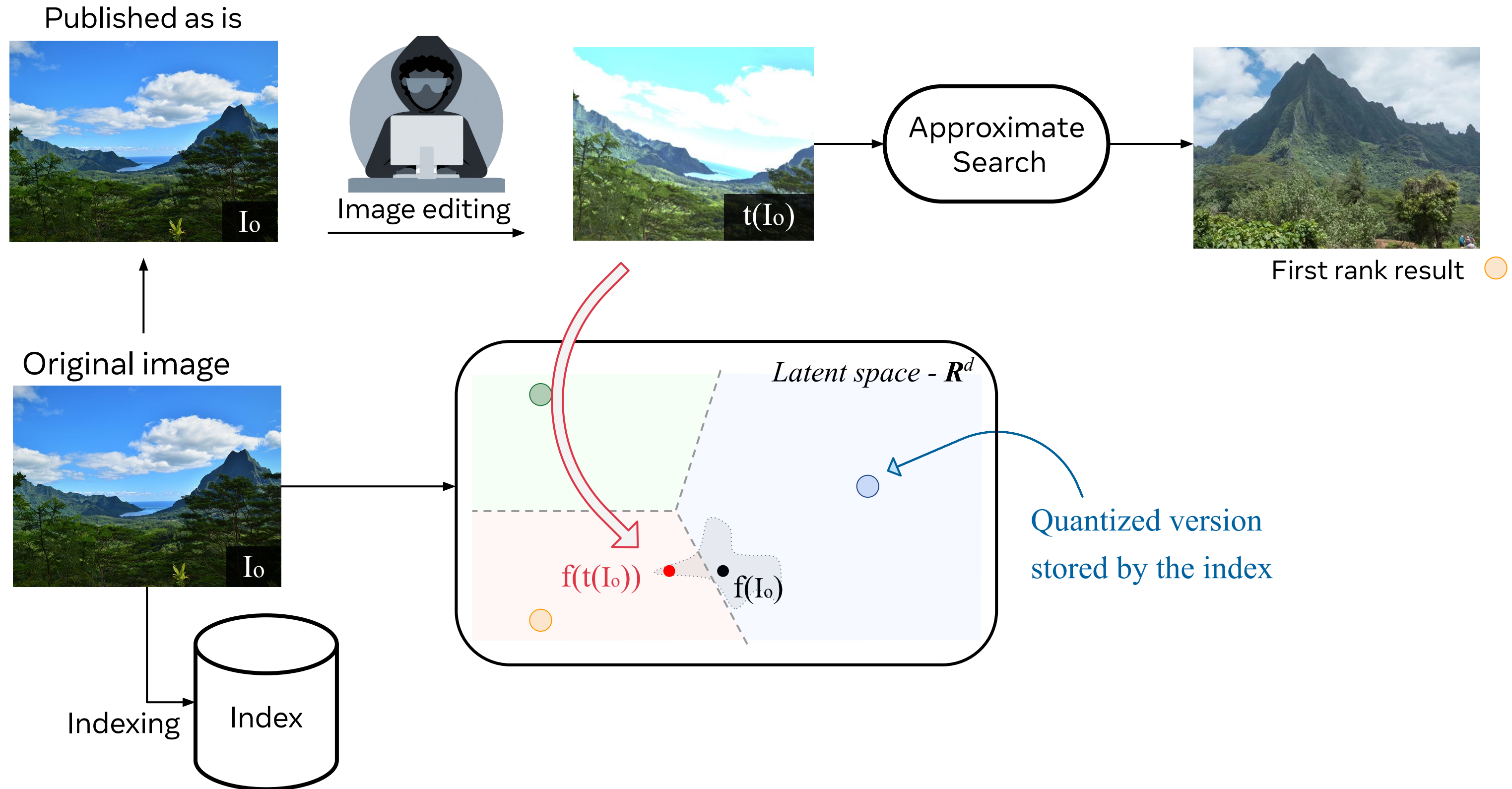
IVF (Inverted File)



Search only in some subsets

Here → **red cell**

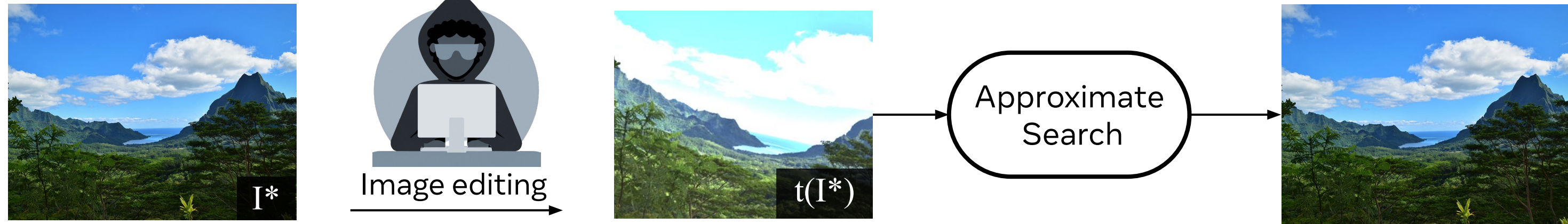
# Presentation of the Problem



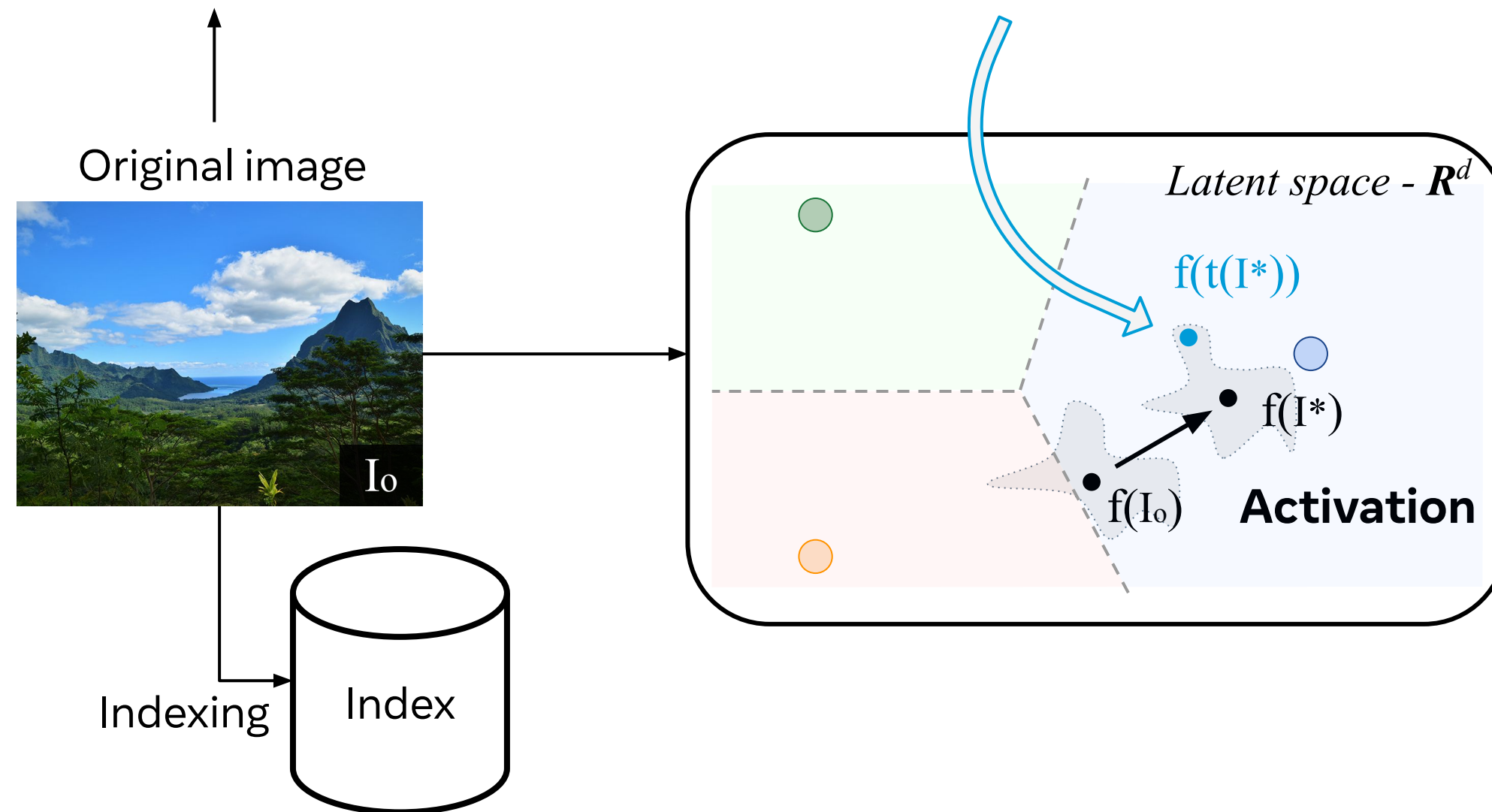


# Presentation of the Problem → Our Improvement

Activated & Published



First rank result

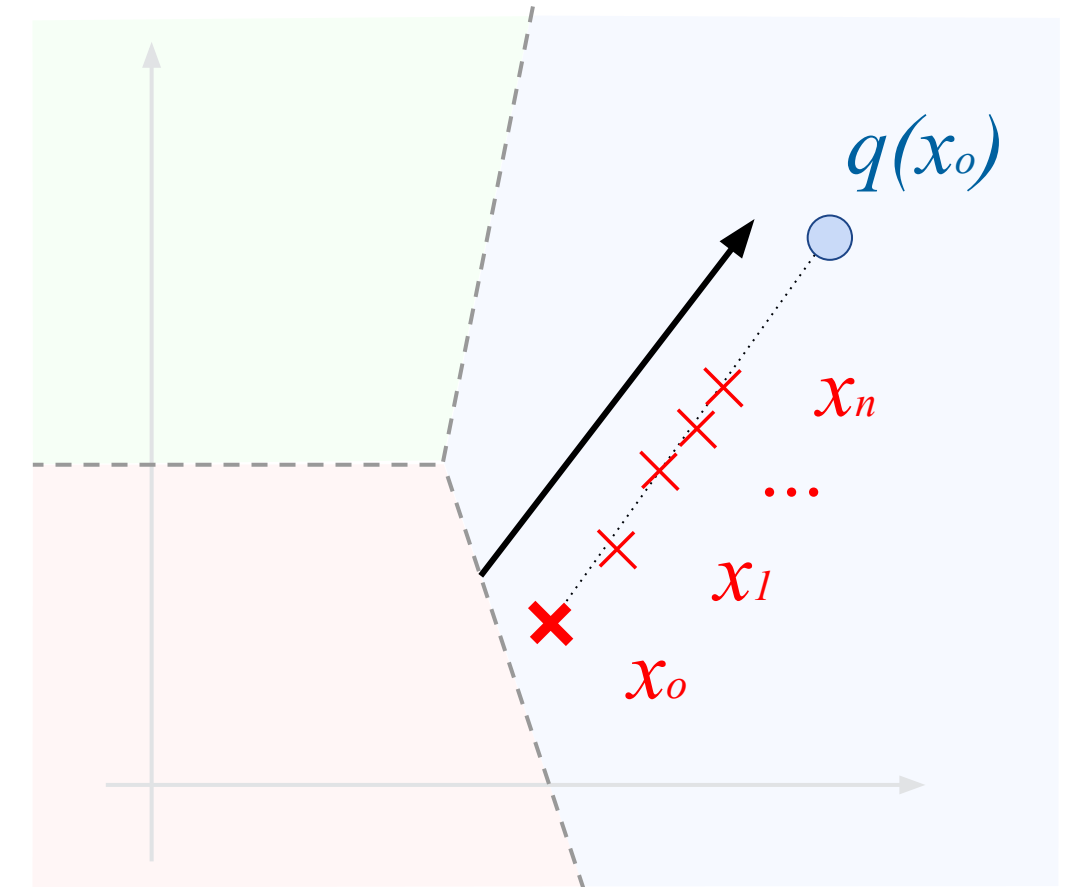
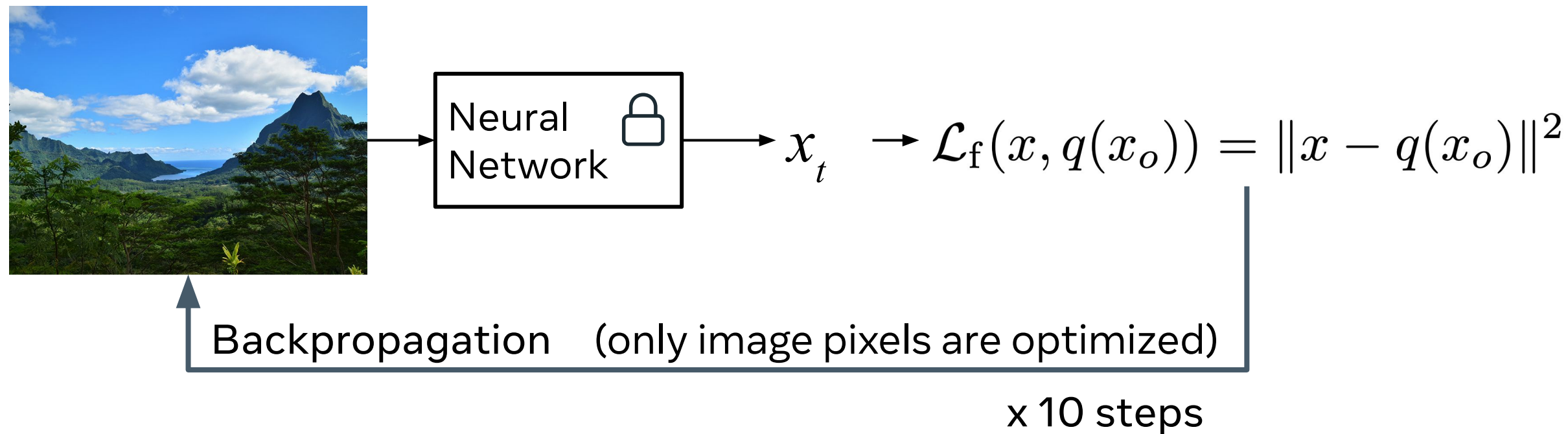


# Method

# Image Optimization

Akin to adversarial attacks [1]

Ex. for PQ:



Effect of the optimization on the latent representation

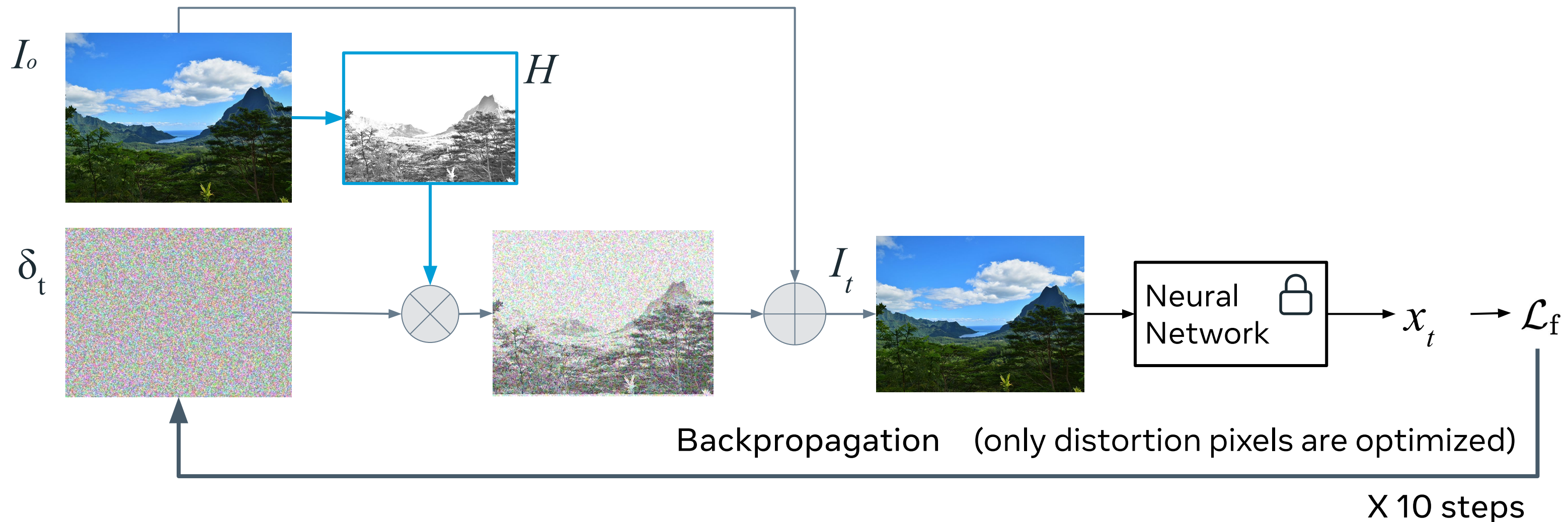
Goal: decrease the quantization error with fixed  $q(x_o) : \|x - q(x_o)\|$

# Perceptual Attenuation (Hand-crafted)

Impose perceptual constraints on activated image?

→ optimize  $\delta$  (not  $I$ )       $I = I_o + \alpha \cdot H_{\text{JND}}(I_o) \odot \tanh(\delta)$

Perceptual model: for each pixel, how much distortion is noticeable for the human eye [2]





# Results

# Experimental Setup

**Dataset** - DISC 2021 (NeurIPS Image Similarity Challenge Dataset) [3]

- 1M reference images,
- Of which 10k queries: edited versions of reference images

**Experiment** - Index 1M reference images, then look the nearest neighbors of query

## Metrics

Recall 1@1: Proportion of 10k queries where the 1<sup>st</sup> rank result is the original image

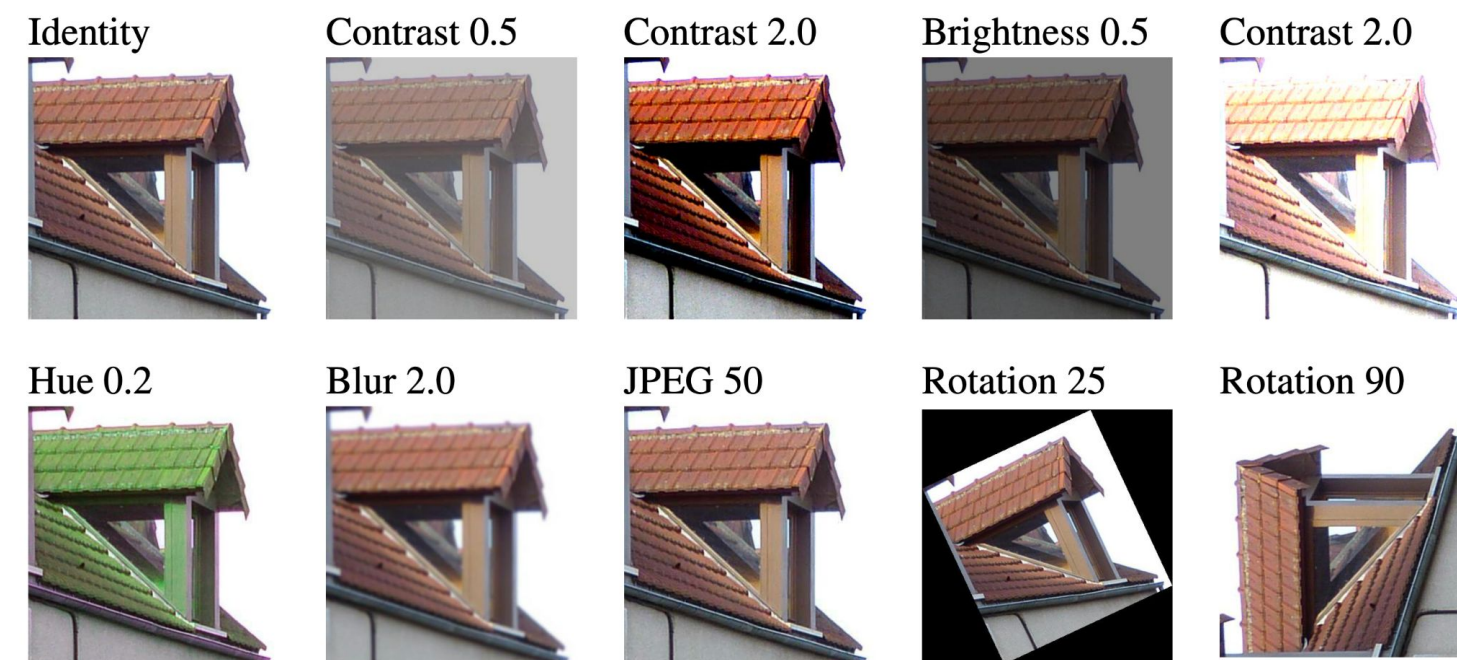
## Parameters

Neural net. extractor: ResNet50 trained with SSCD [4]

*[3. Douze et al. The 2021 image similarity dataset and challenge. Arxiv 2021]*

# R@1 - Retrieval Results

Different transformations are applied to images before search:



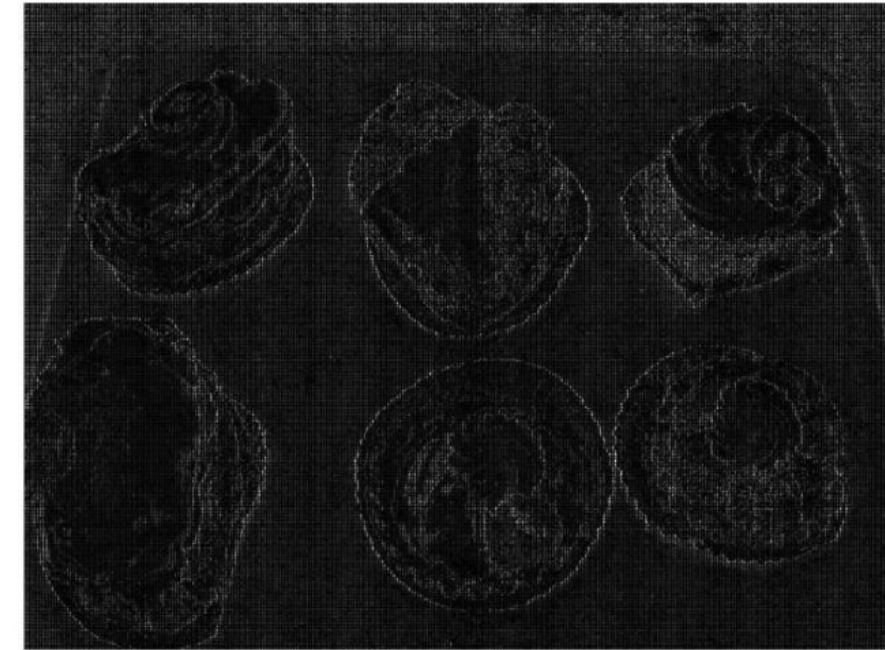
	Search (ms)	Bytes/vector	Activated	Identity	Contr. 0.5	Contr. 2.0	Bright. 0.5	Bright. 2.0	Hue 0.2	Blur 2.0	JPEG 50	Rot. 25	Rot. 90	Crop 0.5	Resi. 0.5	Meme	Random	Avg.
No index	252	2048	✗	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99
IVF-PQ	0.38	8	✗	1.00	0.73	0.39	0.73	0.28	0.62	0.48	0.72	0.07	0.14	0.14	0.72	0.14	0.13	0.45
			✓	1.00	1.00	0.96	1.00	0.92	1.00	0.96	0.99	0.10	0.50	0.29	1.00	0.43	0.32	0.75

Index	Search time	R@1 avg.	
		Passive	Activated
IVF 1024	0.32 ms	0.66	<b>0.89</b>
PCA64, LSH	0.99 ms	0.78	<b>0.86</b>
OPQ 8x8	5.71 ms	0.92	<b>0.94</b>



# Qualitative Results

Avg over 10k images: PSNR=  $43.8 \pm 2.2$  dB - SSIM=  $0.98 \pm 0.01$  -  $\|I - I'\|_{\infty} = 14.5 \pm 1.2$

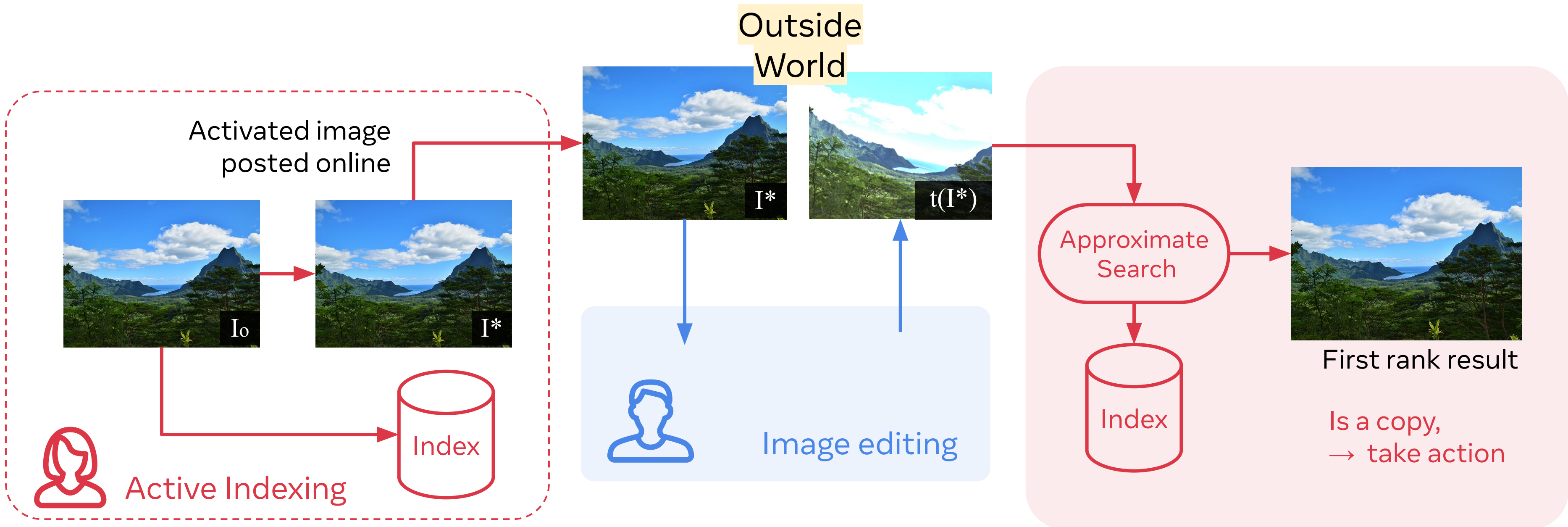




# Conclusion

# Use Case

## Copy Detection of Protected Images



Ex: Shutterstock, Getty, DALL•E, etc.

# Conclusion

## Key takeaways

- Image copy detection: improved if images can be **modified before their release**
- Image optimization scheme: **reduce impact of approximate similarity search**

## Limitations

- **Limited to use-cases** where images can be modified before their release
- **Not agnostic** to index and feature extractor

# Conclusion

## Key takeaways

- Image copy detection: improved if images can be **modified before their release**
- Image optimization scheme: **reduce impact of approximate similarity search**

## Limitations

- **Limited to use-cases** where images can be modified before their release
- **Not agnostic** to index and feature extractor

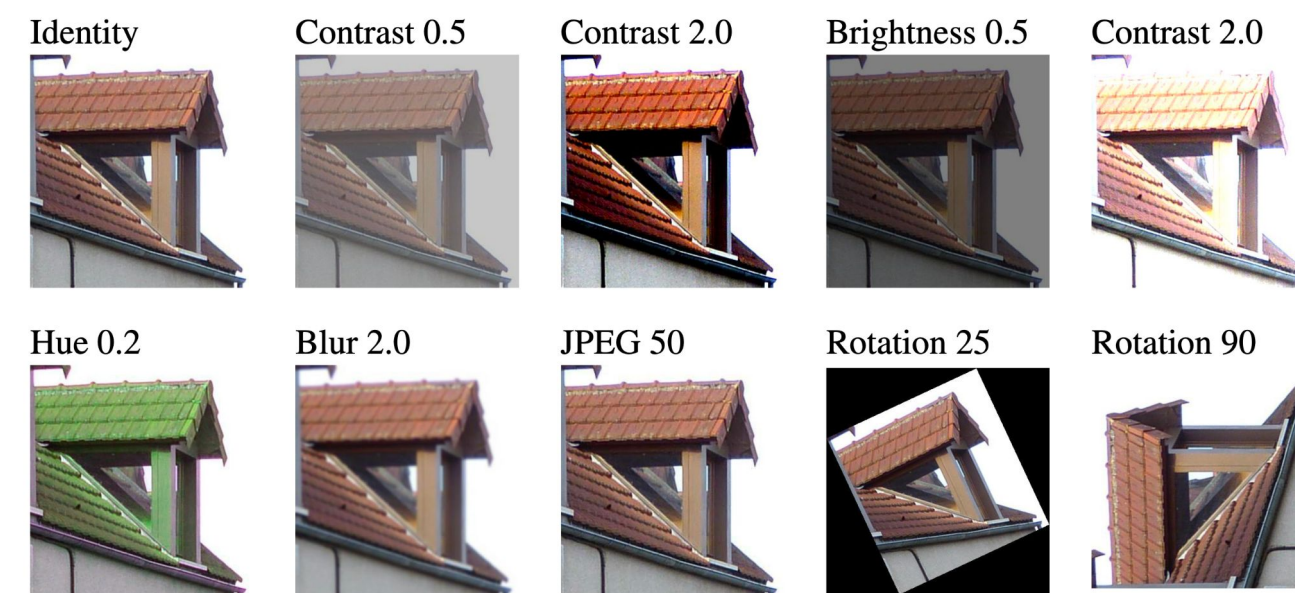
**Thanks for your attention!**



# Appendix

# R@1 - Retrieval Results

Different transfo. are applied to images before search:



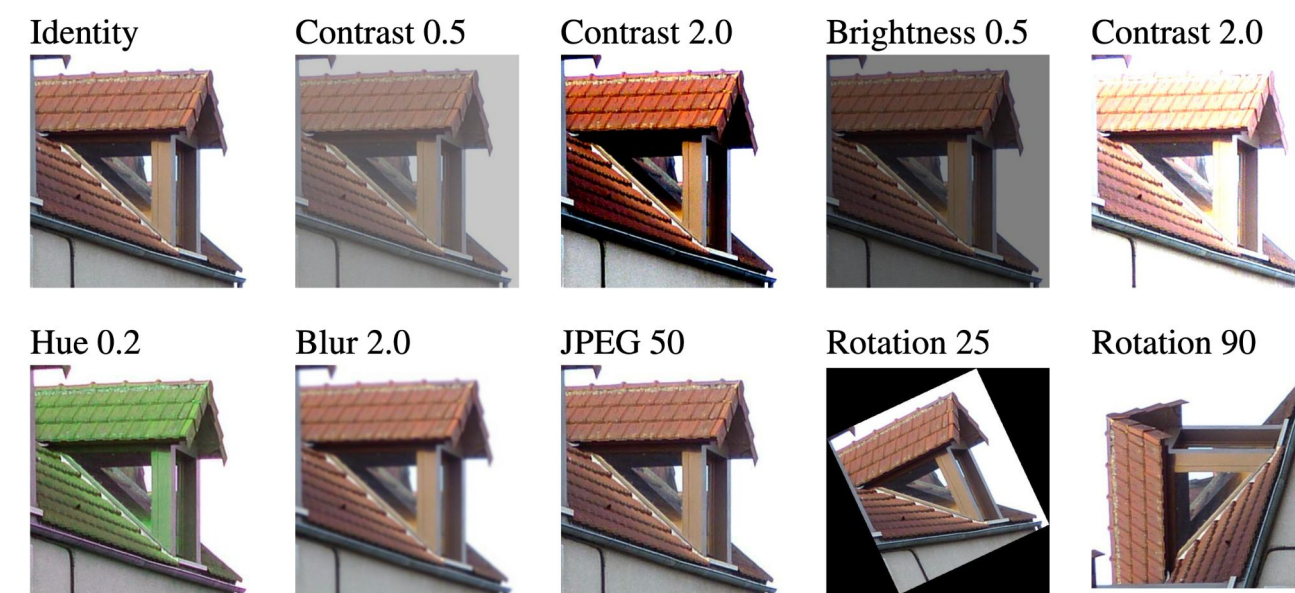
Search (ms) Bytes/vector Activated Identity Contr. 0.5 Contr. 2.0 Bright. 0.5 Bright. 2.0 Hue 0.2 Blur 2.0 JPEG 50 Rot. 25 Rot. 90 Crop 0.5 Resi. 0.5 Meme Random Avg.

No index	252	2048	✗	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99
IVF-PQ	0.38	8	✗	1.00	0.73	0.39	0.73	0.28	0.62	0.48	0.72	0.07	0.14	0.14	0.72	0.14	0.13	0.45	
			✓	1.00	1.00	0.96	1.00	0.92	1.00	0.96	0.99	0.10	0.50	0.29	1.00	0.43	0.32	0.75	
IVF-PQ <sup>16</sup>	0.42	8	✗	1.00	1.00	0.90	1.00	0.78	0.99	0.95	0.99	0.35	0.57	0.57	1.00	0.56	0.39	0.79	
			✓	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.43	0.88	0.75	1.00	0.84	0.50	0.88	



# R@1 - Retrieval Results

Different transfo. are applied to images before search:



	Search (ms)	Bytes/vector	Activated	Identity	Contr. 0.5	Contr. 2.0	Bright. 0.5	Bright. 2.0	Hue 0.2	Blur 2.0	JPEG 50	Rot. 25	Rot. 90	Crop 0.5	Resi. 0.5	Meme	Random	Avg.
--	-------------	--------------	-----------	----------	------------	------------	-------------	-------------	---------	----------	---------	---------	---------	----------	-----------	------	--------	------

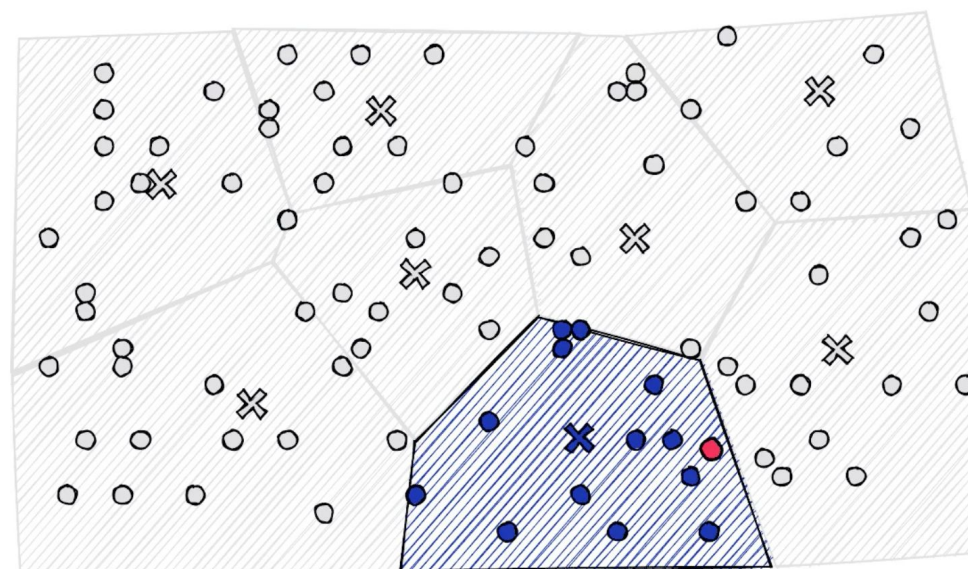
No index	252	2048	✗	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99
IVF-PQ	0.38	8	✗	1.00	0.73	0.39	0.73	0.28	0.62	0.48	0.72	0.07	0.14	0.14	0.72	0.14	0.13	0.45
			✓	1.00	1.00	0.96	1.00	0.92	1.00	0.96	0.99	0.10	0.50	0.29	1.00	0.43	0.32	0.75
IVF-PQ <sup>16</sup>	0.42	8	✗	1.00	1.00	0.90	1.00	0.78	0.99	0.95	0.99	0.35	0.57	0.57	1.00	0.56	0.39	0.79
			✓	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.43	0.88	0.75	1.00	0.84	0.50	0.88



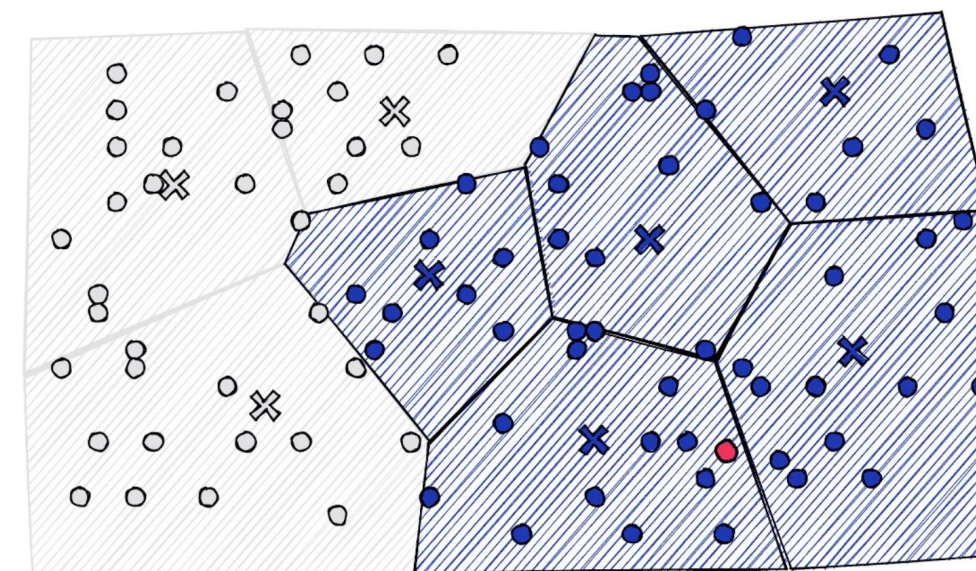
# R@1 - Retrieval Results

IVFPQ vs. IVFPQ<sup>16</sup>

Active 1 probe  $\approx$  Passive 16 probes



$nprobe = 1$



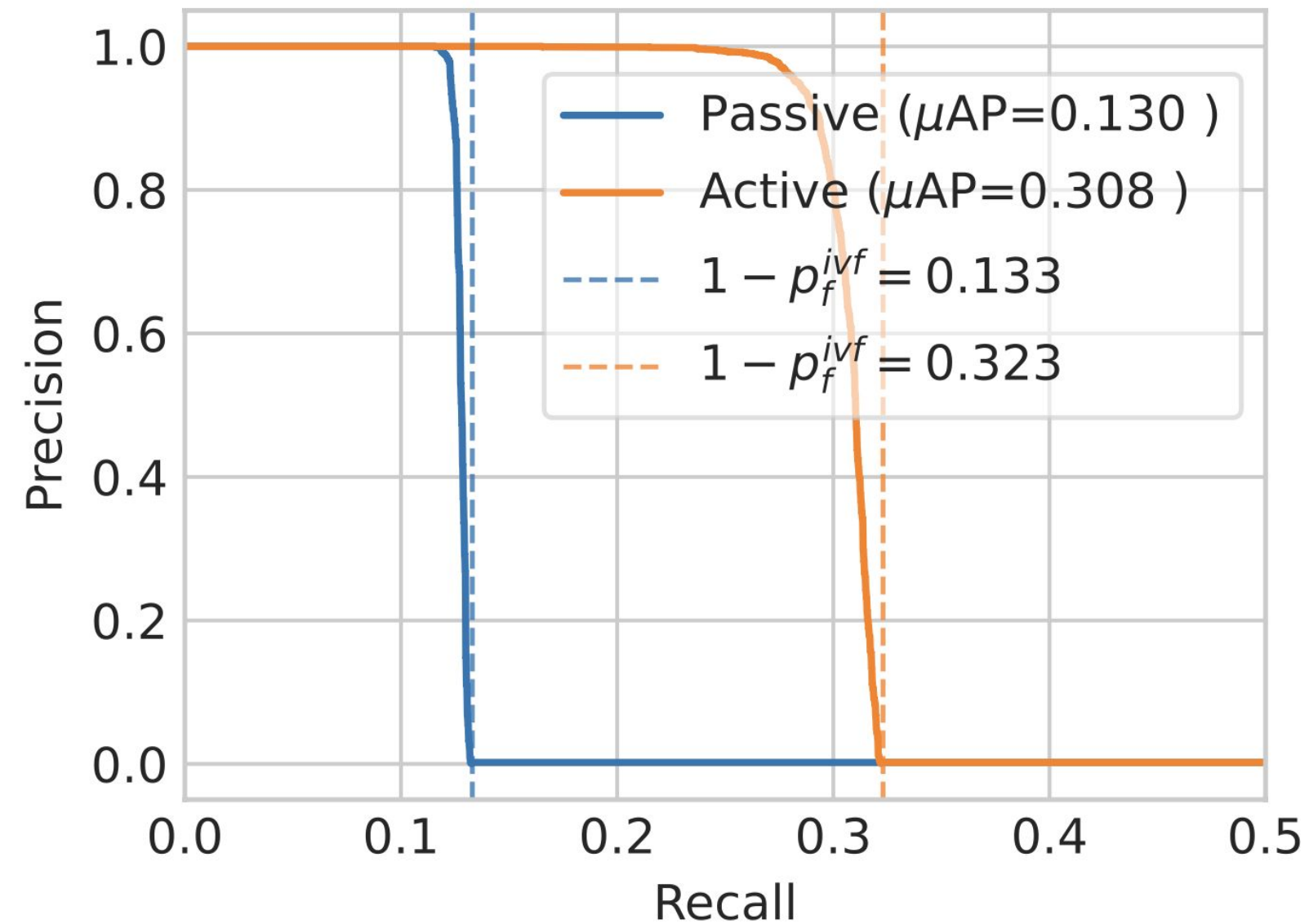
$nprobe = 5$

	Search (ms)	Bytes/vector	Activated	Identity	Contr. 0.5	Contr. 2.0	Bright. 0.5	Bright. 2.0	Hue 0.2	Blur 2.0	JPEG 50	Rot. 25	Rot. 90	Crop 0.5	Resi. 0.5	Meme	Random	Avg.
No index	252	2048	✗	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.99
IVF-PQ	0.38	8	✗	1.00	0.73	0.39	0.73	0.28	0.62	0.48	0.72	0.07	0.14	0.14	0.72	0.14	0.13	0.45
			✓	1.00	1.00	0.96	1.00	0.92	1.00	0.96	0.99	0.10	0.50	0.29	1.00	0.43	0.32	0.75
IVF-PQ <sup>16</sup>	0.42	8	✗	1.00	1.00	0.90	1.00	0.78	0.99	0.95	0.99	0.35	0.57	0.57	1.00	0.56	0.39	0.79
			✓	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.43	0.88	0.75	1.00	0.84	0.50	0.88



# PRC - Copy Detection Results

x2 in Area under PRC

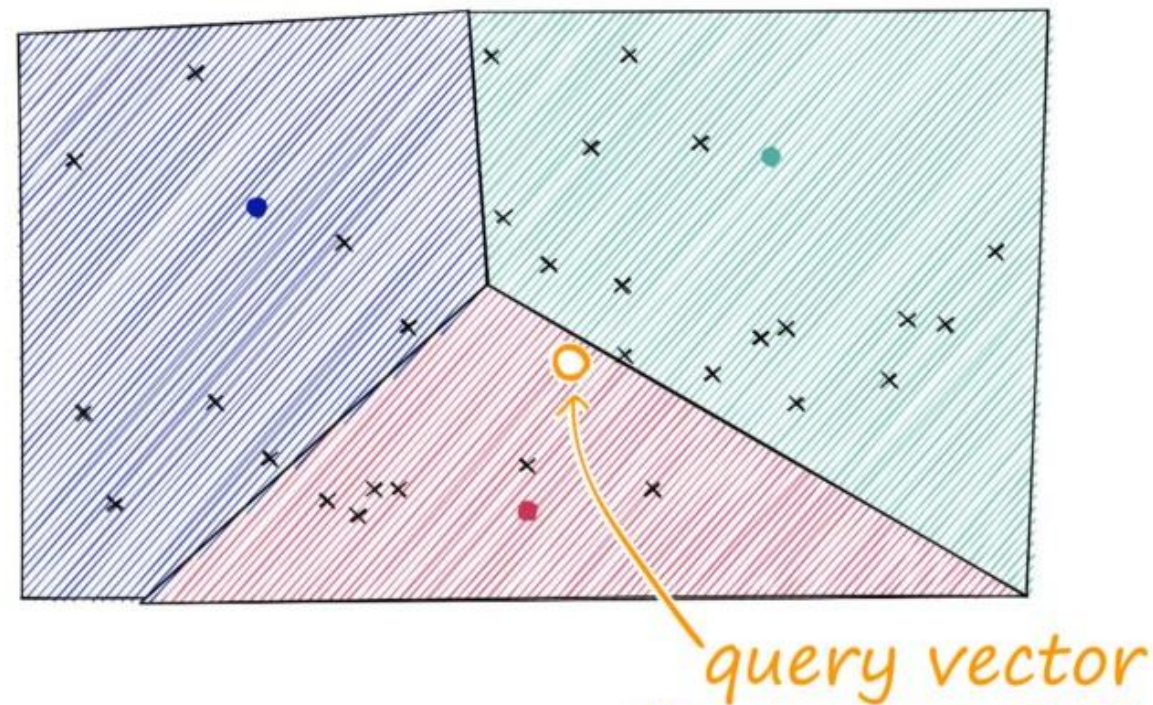


# Some Indexing Methods

## IVF

(Inverted File)

→ *Space partitioning*



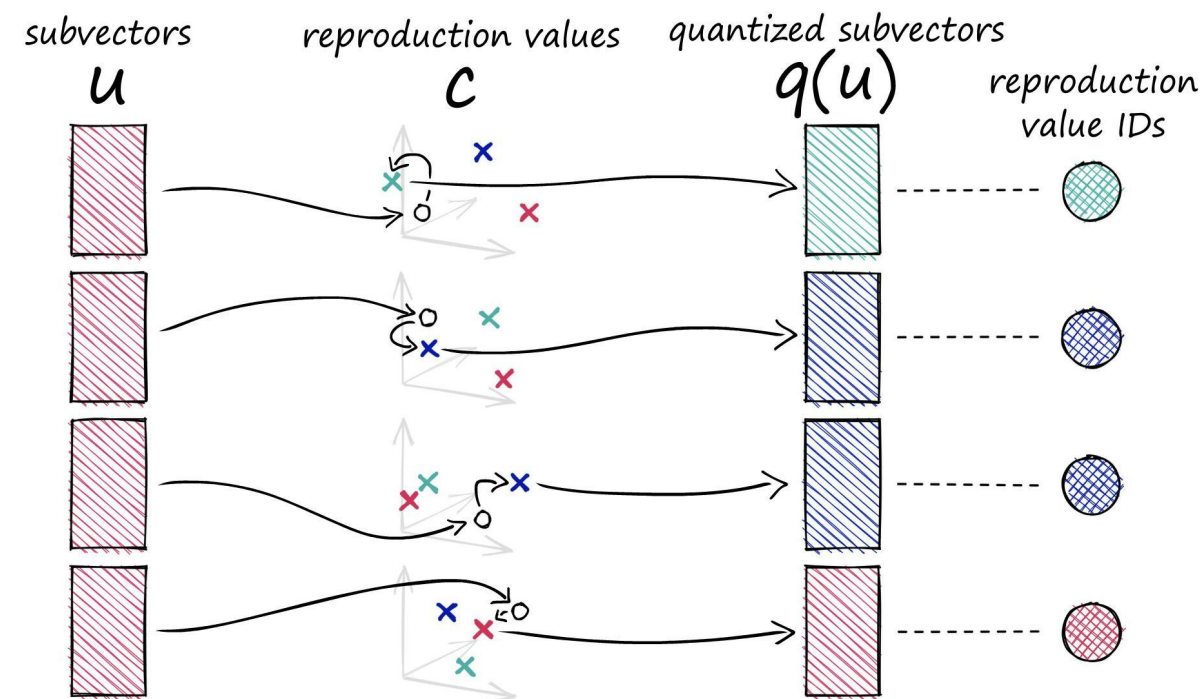
Search only in some cells

Here → **red cell**

## PQ

(Product Quantization)

→ *Vector compression*



Produces a list of centroids:

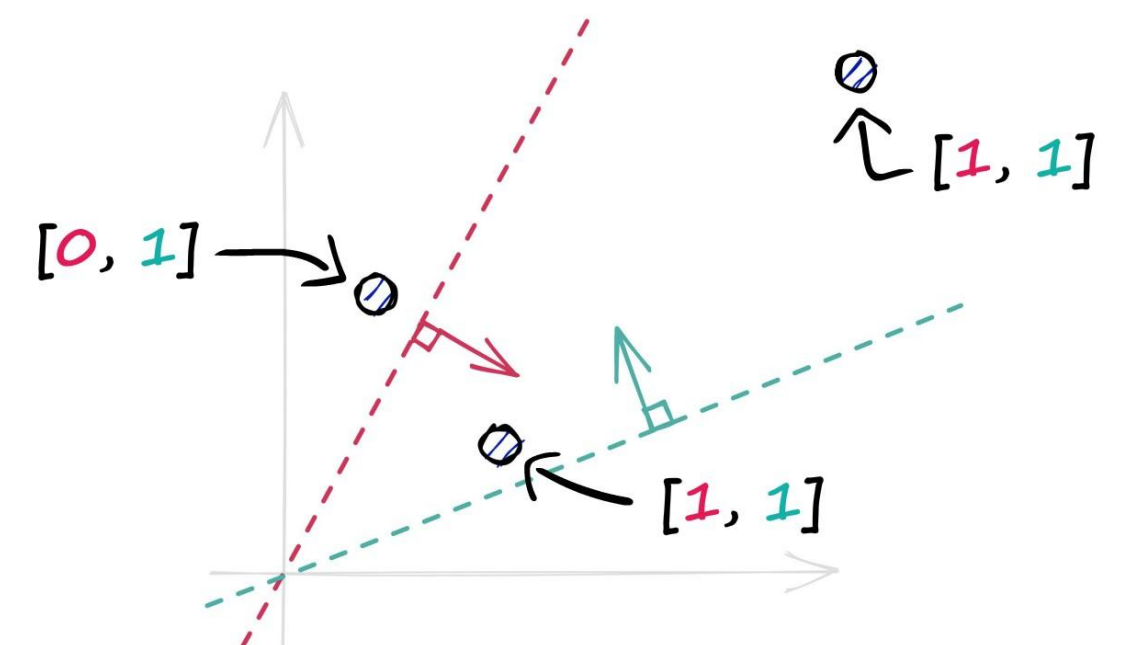
$x$  is represented by

$$[c_1, c_2, \dots, c_m] \text{ with } c_i \in [0, 2^k]$$

## LSH

(Locality Sensitive Hashing)

→ *Both (depends on context)*



Produces binary codes ( $[0, 1]$ ):

Left side of  $/$  → 0

Right side of  $-$  → 1



# $L^\infty$ bound VS. perceptual model



(a)  $\ell_\infty = 4$ , PSNR = 36.4 dB, SSIM = 0.91



(b)  $\ell_\infty = 23$ , PSNR = 34.4 dB, SSIM = 0.94



# Time

	Activation	Identity	Contr. 0.5	Contr. 2.0	Bright. 0.5	Bright. 2.0	Hue 0.2	Blur 2.0	JPEG 50	Rot. 25	Rot. 90	Crop 0.5	Resi. 0.5	Meme	Random	Avg.
Passive	-	1.00	0.73	0.39	0.73	0.28	0.62	0.48	0.72	0.07	0.14	0.14	0.72	0.14	0.13	0.45
Adam,lr=1 - 10 steps	39.8 ms/img	1.00	1.00	0.96	1.00	0.92	1.00	0.96	0.99	0.10	0.50	0.29	1.00	0.43	0.32	0.75
Adam,lr=10 - 1 step	4.3 ms/img	0.99	0.99	0.92	0.99	0.84	0.99	0.95	0.99	0.10	0.39	0.25	0.99	0.36	0.27	0.72