



Outline

- ▶ GAN for **unsupervised** learning of image **binary descriptors**.
- ▶ **Binary representation** of the GAN's **discriminator**.
- ▶ **Two regularizations** to improve binary representation.
- ▶ First transfers distances from **high-dimensional** to **low-dimensional** space.
- ▶ The second one increases the **diversity** of binary vectors.
- ▶ Evaluation on **two** applications: **image matching** and **retrieval**.
- ▶ Our approach significantly **outperforms state-of-the-art methods**.

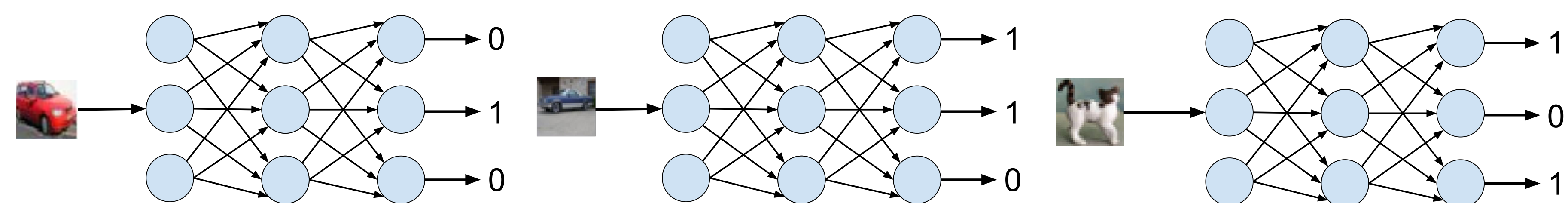
Contributions

- ▶ Compact **binary** image **descriptors** learned in an **unsupervised manner**.
- ▶ Two regularization methods to achieve **binary** embeddings on discriminator.
- ▶ SOTA results in two tasks: **image retrieval** and **matching**.

Learning binary descriptors

Binary descriptors are feature embeddings that are: **low-dimensional**, **quantized**.

- ▶ **Past**: hand-crafted feature descriptors have been used for representing images.
- ▶ **Present**: deep neural networks are used to learn binary features.
- ▶ **Ours**: using GAN's generator to learn binary codes.



GAN model

We consider two networks: **generator** (G), **discriminator** (D).

The **generator** is trained using **feature matching**:

$$||\mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} \mathbf{f}(\mathbf{x}) - \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} \mathbf{f}(G(\mathbf{z}))||_2^2.$$

The **discriminator** is trained using typical criterion:

$$L_D = -\mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] - \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))].$$

BinGAN model

- ▶ The **generator** is trained using **feature matching**.
- ▶ The **discriminator** is trained with the aggregated criterion:

Discriminator Loss

$$L_D = -\mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] - \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

BinGAN Loss

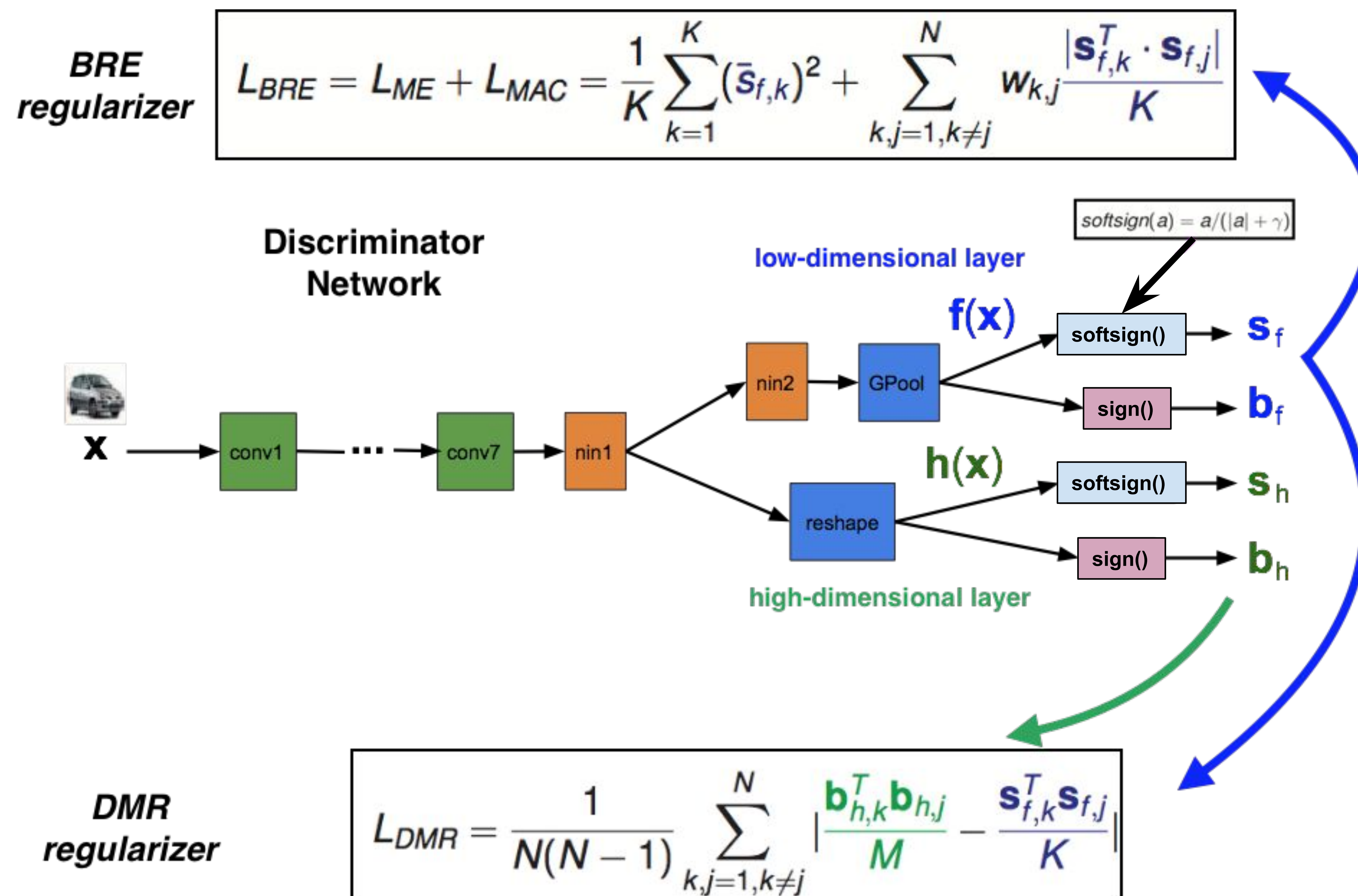
$$L = L_D + \lambda_{BRE} \cdot L_{BRE} + \lambda_{DMR} \cdot L_{DMR}.$$

Binary Representation Entropy (BRE)

Distance Matching Regularizer (DMR)

Training BinGAN

- ▶ **Distance Matching Regularizer (DMR)** is responsible for matching distances from **high-dimensional** ($\mathbf{h}(\cdot)$) to **low-dimensional** ($\mathbf{f}(\cdot)$) space.
- ▶ **Adjusted Binarization Representation Entropy (BRE) Regularizer** is increasing the diversity of binary vectors in the low-dimensional layer:
 - ▶ L_{ME} is responsible for **enforcing uniform distribution** of $\mathbf{s}_f$.
 - ▶ L_{MAC} is responsible for minimizing the **weighted correlation**.

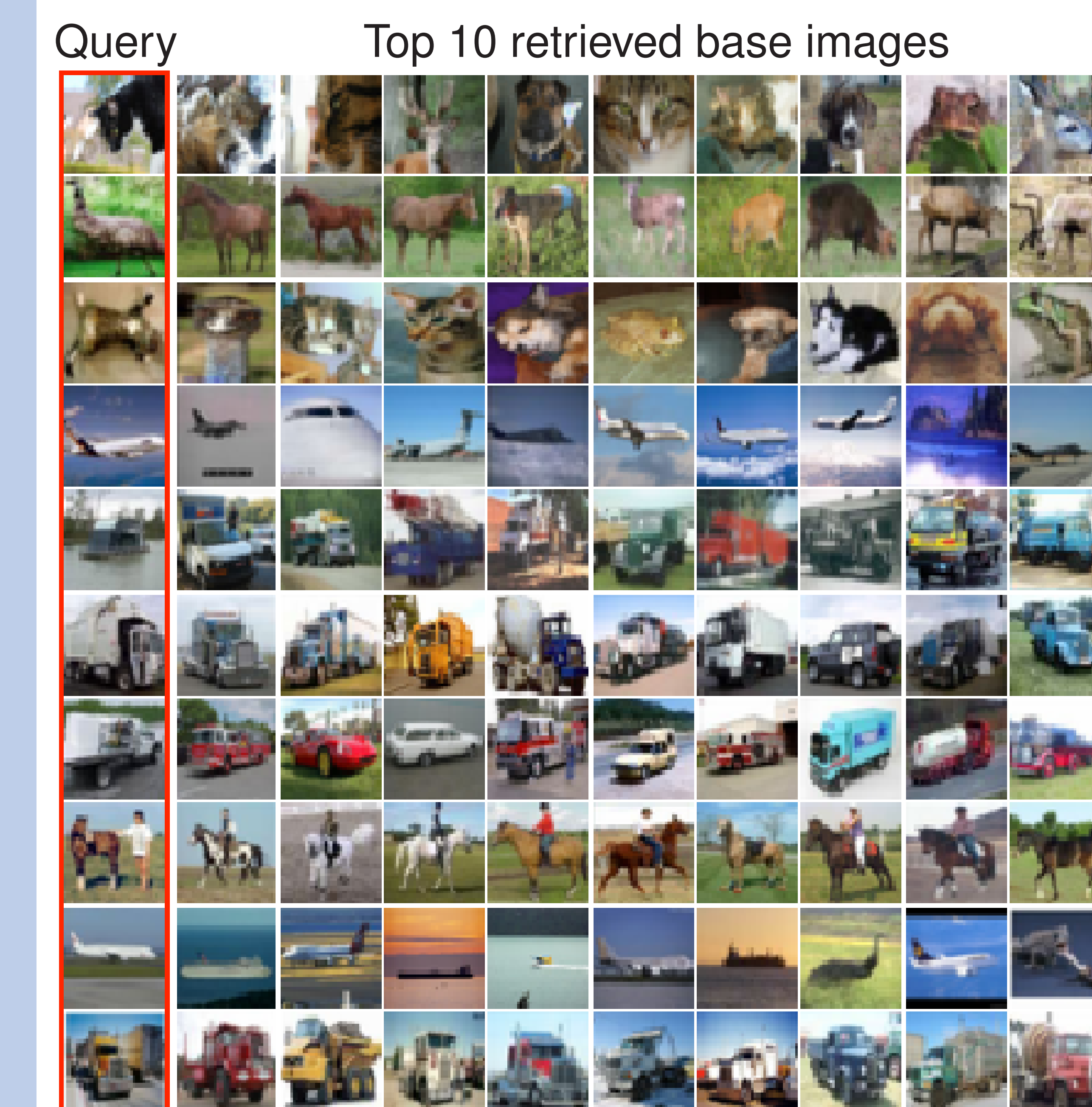


- ▶ The functions $\mathbf{f}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ denote the **low** and **high-dimensional** intermediate layers of discriminator with the numbers of hidden units equal to K and M , respectively.
- ▶ The corresponding **binary codes** (using $\text{sign}(\cdot)$) are denoted by $\mathbf{b}_f$ and $\mathbf{b}_g$.
- ▶ The corresponding **quantized codes** (using $\text{softsign}(\cdot)$) are denoted by $\mathbf{s}_f$ and $\mathbf{s}_g$.

Results - image retrieval (Cifar10)

Mean Average Precision (mAP) - top 1000.

Method	16 bit	32 bit	64 bit
KHM	13.59	13.93	14.46
SphH	13.98	14.58	15.38
SpeH	12.55	12.42	12.56
SH	12.95	14.09	13.89
PCAH	12.91	12.60	12.10
LSH	12.55	13.76	15.07
PCA-ITQ	15.67	16.20	16.64
DH	16.17	16.62	16.96
DeepBit	19.43	24.86	27.73
DBD-MQ	21.53	26.50	31.85
BinGAN	30.05	34.65	36.77

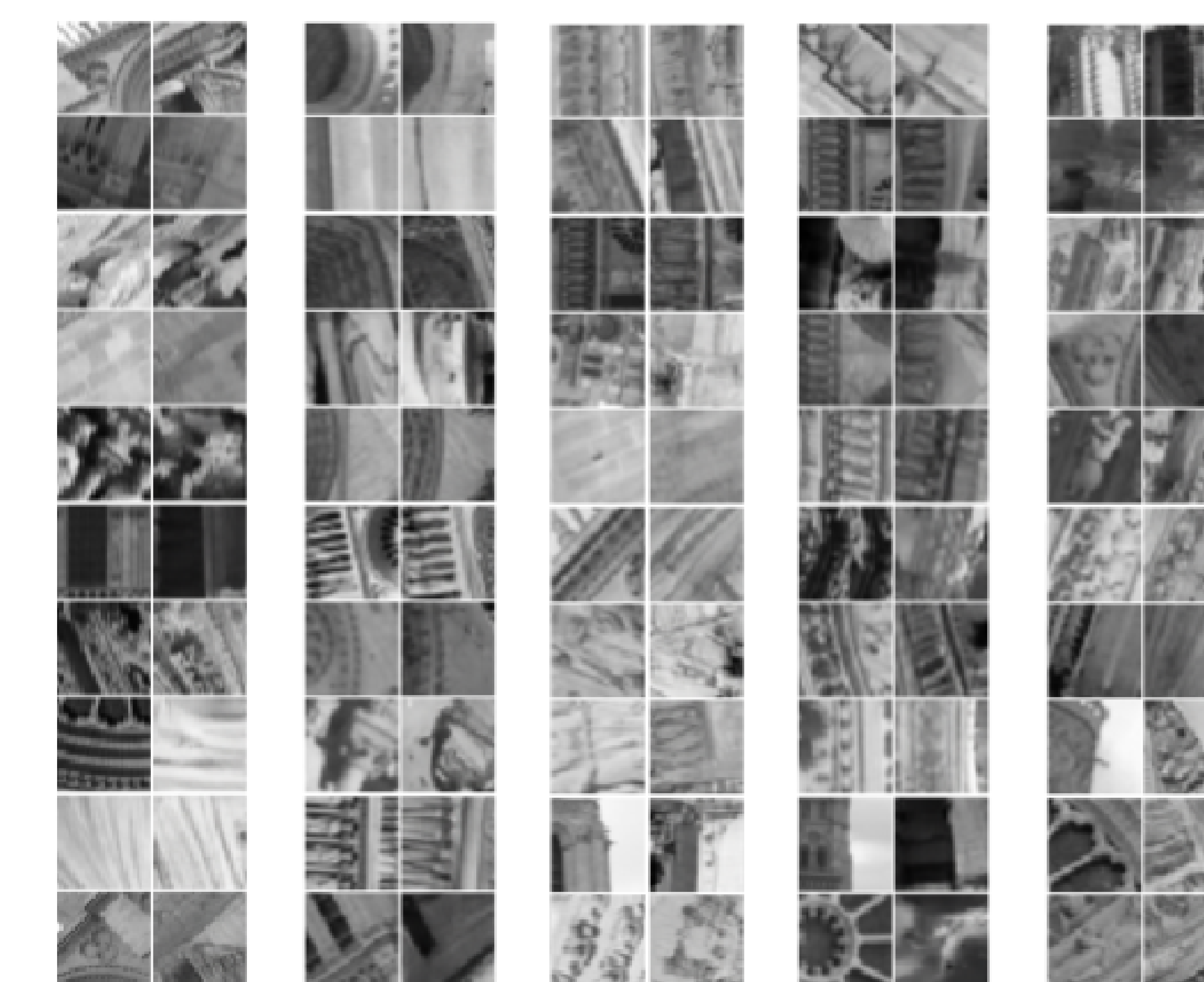


Results - image matching (Brown)

The **false positive** rates at 95% true positives (FPR@95%).

Train Test	Yosemite		Notre Dame		Liberty		Average FPR@95%
	Notre Dame	Liberty	Yosemite	Liberty	Notre Dame	Yosemite	
Supervised							
LDAHash (16 bytes)	51.58	49.66	52.95	49.66	51.58	52.95	51.40
D-BRIEF (4 bytes)	43.96	53.39	46.22	51.30	43.10	47.29	47.54
BinBoost (8 bytes)	14.54	21.67	18.96	20.49	16.90	22.88	19.24
RFD (50-70 bytes)	11.68	19.40	14.50	19.35	13.23	16.99	15.86
Binary L2-Net (32 bytes)	2.51	6.65	4.04	4.01	1.9	5.61	4.12
Unsupervised							
SIFT (128 bytes)	28.09	36.27	29.15	36.27	28.09	29.15	31.17
BRISK (64 bytes)	74.88	79.36	73.21	79.36	74.88	73.21	75.81
BRIEF (32 bytes)	54.57	59.15	54.96	59.15	54.57	54.96	56.23
DeepBit (32 bytes)	29.60	34.41	63.68	32.06	26.66	57.61	40.67
DBD-MQ (32 bytes)	27.20	33.11	57.24	31.10	25.78	57.15	38.59
BinGAN (32 bytes)	16.88	26.08	40.80	25.76	27.84	47.64	30.76

Image matching - qualitative analysis



True patches (left) and the closest generated (right).