Higher-order Statistical Modeling based Deep CNNs （Part-I）

Classical Methods & From Shallow to Deep

Qilong Wang

2018-11-23

Context

1. Higher-order Statistics in Bag-of-Visual-Words (BoVW)
2. Higher-order Statistics in Codebookless Model (CLM)
3. Bag-of-Visual-Words *vs.* Codebookless Model
4. Higher-order Statistical Models Meet Deep Features

Higher-order Statistics

Higher-order Statistics

1st-order Vector	2nd-order Matrix	3rd-order Tensor	……
$\sum \mathbf{X}$	$\mathbf{X}^T\mathbf{X}$	$\mathbf{X}\otimes\mathbf{X}\otimes\mathbf{X}$

Context

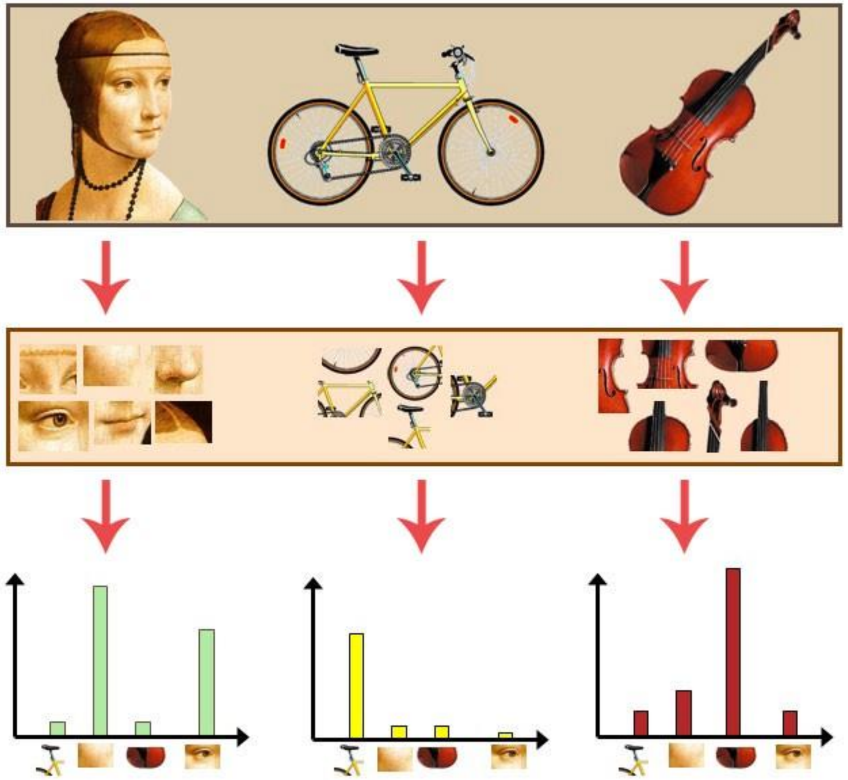
1
- **Higher-order Statistics in Bag-of-Visual-Words (BoVW)**

2
- **Higher-order Statistics in Codebookless Model (CLM)**

3
- **Bag-of-Visual-Words *vs.* Codebookless Model**

4
- **Higher-order Statistical Models Meet Deep Features**

Bag-of-Visual-Words (BoVW)



~75,300

2003 - 2012

[1] J. Sivic and A. Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. ICCV, 2003. ***(cited by 6391)***
[2] C. Dance, J. Willamowski et al. Visual categorization with bags of keypoints. ECCV Workshop, 2004. ***(cited by 4767)***

BoVW – Comparison

VOC07
[BoW + VQ]

VOC08
[BoW + VQ]

VOC09
BoW + [VQ+ LLC + SV]

VOC10&11
[BoW + Context]

VOC12
[BoW + VQ+ LLC + FV]

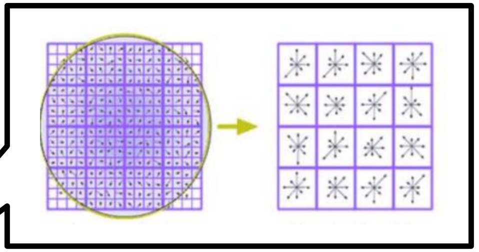
All winners (classification) based on BoVW!

The PASCAL Visual Object Classes Homepage

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

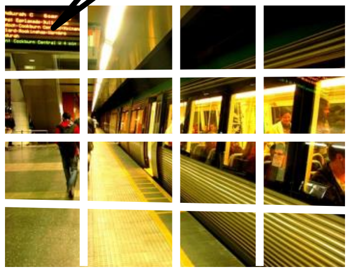
Bag-of-Visual-Words (BoVW)

SIFT [IJCV03]



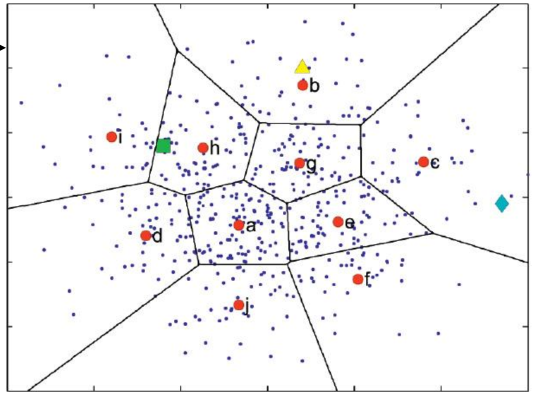
$$h(i) = \frac{1}{N}\sum_{n=1}^{N}\begin{cases}1 & i = \underset{k}{\operatorname{argmin}}\, D(w_k, x_n)\\ 0 & \text{otherwise}\end{cases}$$

Image

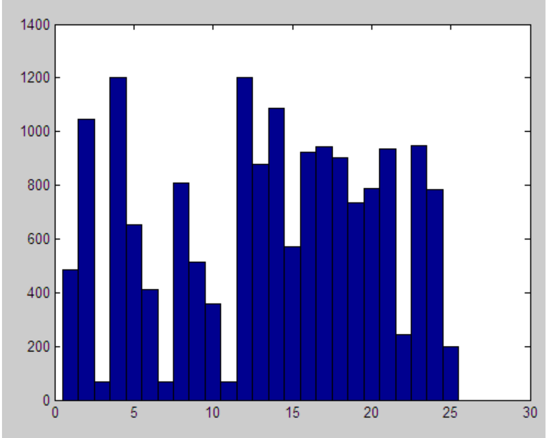


Local Features

Centers of K-means



Dictionary



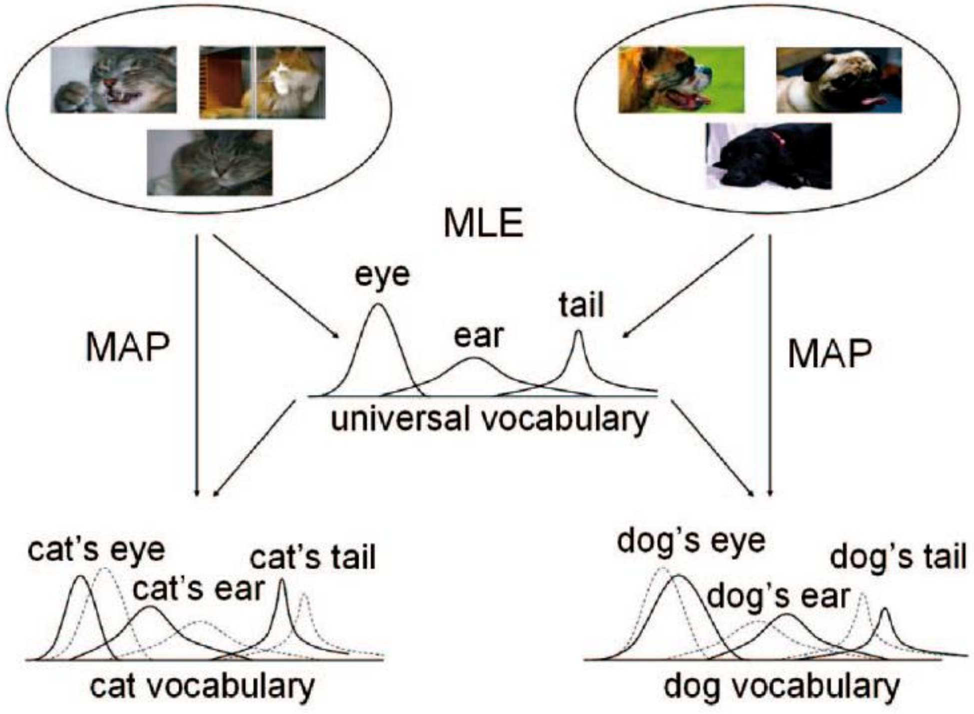
Histogram

Training Images

0th-order coding

[1] J. Sivic and A. Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. ICCV, 2003. ***(cited by 6391)***
[2] C. Dance, J. Willamowski et al. Visual categorization with bags of keypoints. ECCV Workshop, 2004. ***(cited by 4767)***

BoVW – Soft Coding



Each atom is a Gaussian.

$$h(i) = \frac{1}{N}\sum_{n=1}^{N}\frac{K_\sigma(D(w_i, x_n))}{\sum_j K_\sigma(D(w_j, x_n))}$$

$$K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right)$$

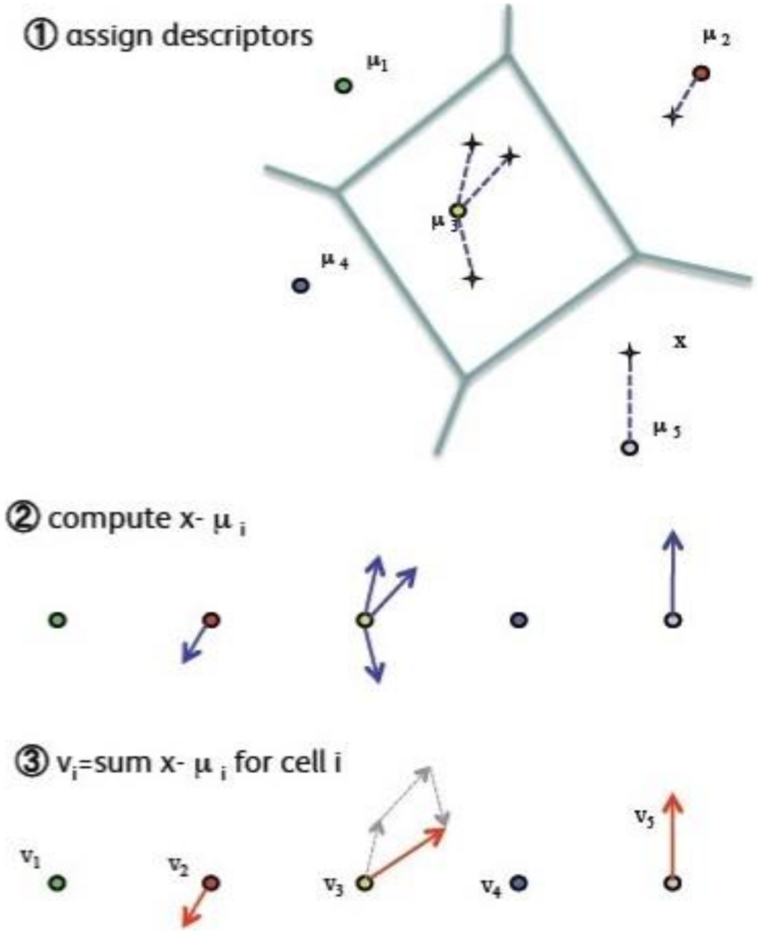
or

$$x \sim p(x|\lambda) = \sum_k \omega_k p(x|q = k, \lambda)$$

Higher-order Dictionary but 0th-order Coding!

[1] Florent Perronnin. Universal and Adapted Vocabularies for Generic Visual Categorization. *TPAMI, 2008*.
[2] Van Gemert, *et al*. Visual Word Ambiguity. *TPAMI*, 2009.

BoVW – Super Vector



1st-order Dictionary &1st-order Coding!

VLAD:

$$v_{i,j} = \sum_{x \text{ such that } \mathrm{NN}(x)=c_i} x_j - c_{i,j}$$

Super Vector (SV) [BoVW + VLAD]:

$$\phi(x) = \left[\underbrace{0,\ldots,0}_{d+1 \text{ dim.}}, \boxed{\underbrace{s, (x-v)^\top}_{d+1 \text{ dim.}}}, \underbrace{0,\ldots,0}_{d+1 \text{ dim.}} \right]^\top$$

[1] Herve J ´ egou *et al*. Aggregating local descriptors into a compact image representation. CVPR, 2010.
[2] Zhou et al. Image Classification using Super-Vector Coding of Local Image Descriptors. ECCV, 2010.

BoVW – Universal GMM

Gaussian Mixture Model as Dictionary

- Adaptive GMM [CVPR, 2008]
- Gaussianized Vector Representation [PRL, 2010]
- Fisher Vector [IJCV, 2013].

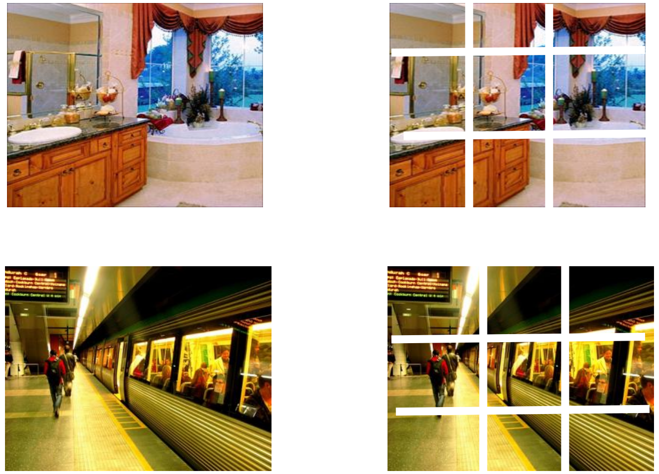
BoVW – Adaptive GMM

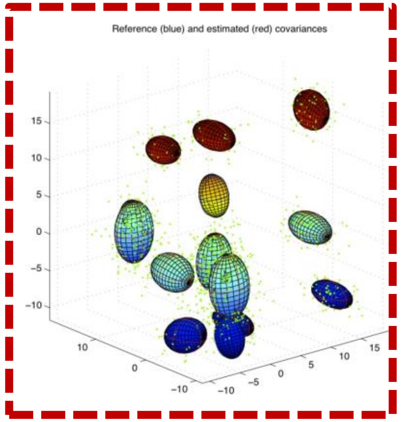
$$
\begin{aligned}
\hat{w}_i^a &= \frac{\sum_{t=1}^{T}\gamma_i(x_t)+\tau}{T+N\times\tau}, \\
\hat{\mu}_i^a &= \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t+\tau\mu_i^u}{\sum_{t=1}^{T}\gamma_i(x_t)+\tau}, \\
\hat{\Sigma}_i^a &= \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t x_t'+\tau[\Sigma_i^u+\mu_i^u\mu_i^{u\prime}]}{\sum_{t=1}^{T}\gamma_i(x_t)+\tau} \\
&\quad -\hat{\mu}_i^a\hat{\mu}_i^{a\prime}.
\end{aligned}
$$

MAP estimation

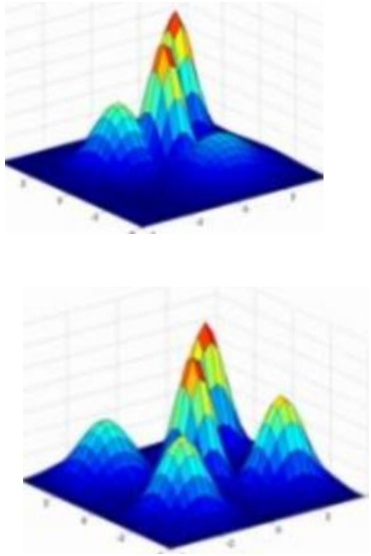
Liu et al. A similarity measure between unordered vector sets with application to image categorization. [CVPR 08]

Unstable & High Cost!





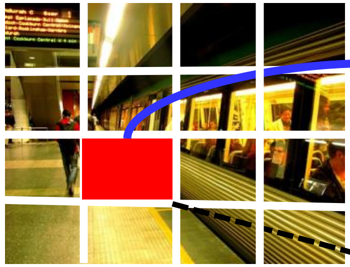
Universal GMM



KL Kernel

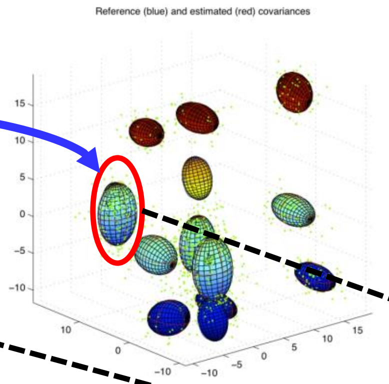
BoVW – Gaussianized Vector

Image



Local Features

NN



Universal GMM

$$\hat{Z} = \begin{pmatrix} \left(\frac{\Sigma_1}{n_1}\right)^{-\frac{1}{2}} Z_1 \\ \boxed{\left(\frac{\Sigma_2}{n_2}\right)^{-\frac{1}{2}} Z_2} \\ \vdots \\ \left(\frac{\Sigma_M}{n_M}\right)^{-\frac{1}{2}} Z_M \end{pmatrix}$$

$$Z_k = \frac{1}{n_k}\sum_{i=1}^{n_k}(Y_{ki} - \mu_k)$$

higher-order Dictionary &1st-order Coding!

Zhou et al. Novel Gaussianized vector representation for improved natural scene categorization. *PRL, 2010*.

BoVW – Fisher Vector

Idea: Representing a random sample *X* with gradients of the distribution

The *steepest descent direction* of $\log p(X \mid \theta)$ in a Riemannian manifold is $\mathbf{I}_\theta^{-1}\nabla_\theta \log p(X \mid \theta)$, which is called *natural gradient*

$$\begin{aligned}\langle X_1, X_2\rangle_\theta &= \left\langle \mathbf{I}_\theta^{-1}\nabla_\theta \log p(X_1 \mid \theta),\ \mathbf{I}_\theta^{-1}\nabla_\theta \log p(X_2 \mid \theta)\right\rangle_\theta \\ &= (\mathbf{I}_\theta^{-1}\nabla_\theta \log p(X_1 \mid \theta))^T \mathbf{I}_\theta \mathbf{I}_\theta^{-1}\nabla_\theta \log p(X_2 \mid \theta) \\ &= \nabla_\theta \log p(X_1 \mid \theta) \mathbf{I}_\theta^{-1}\nabla_\theta \log p(X_2 \mid \theta)\end{aligned}$$

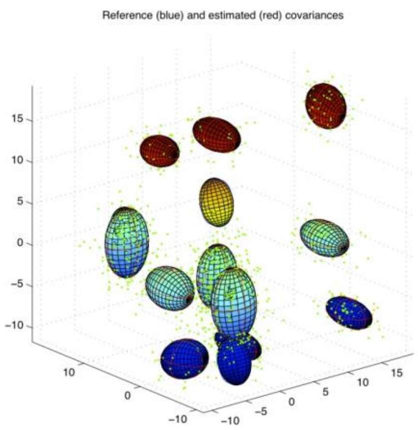
Fisher vector:

$$X \rightarrow \mathbf{I}_\theta^{-1/2}\ \nabla_\theta \log p(X \mid \theta)$$

?

Tommi S. Jaakkola and David Haussler. Exploiting generative models in discriminative classifiers. NIPS, 1998.

BoVW – Fisher Vector



Universal GMM

$$\left[w_k , \boldsymbol{\mu}_k , \boldsymbol{\sigma}_k \right]$$

$$\mathscr{G}^X_{\alpha_k} = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \left(\boxed{\gamma_t(k)} - w_k \right),$$

Weight 0^{th}-order

$$\mathscr{G}^X_{\mu_k} = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \boxed{\gamma_t(k)} \left(\frac{\boxed{x_t} - \mu_k}{\sigma_k} \right),$$

Mean 1^{st}-order

$$\mathscr{G}^X_{\sigma_k} = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \boxed{\gamma_t(k)} \frac{1}{\sqrt{2}} \left[\frac{(\boxed{x_t} - \mu_k)^2}{\sigma_k^2} - 1 \right]$$

Variance 2^{nd}-order

Posterior probability

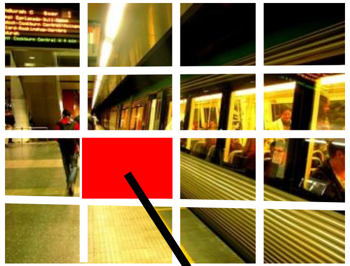
Local features

[1] Florent Perronnin *et al*. Improving the Fisher Kernel for Large-Scale Image Classification. ECCV, 2010.

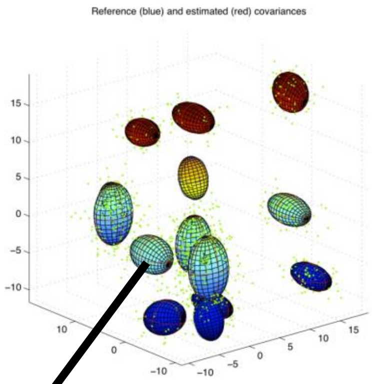
[2] Sánchez *et al*. Image classification with the fisher vector: Theory and practice. IJCV, 2013.

BoVW – Fisher Vector

Image



Local Features

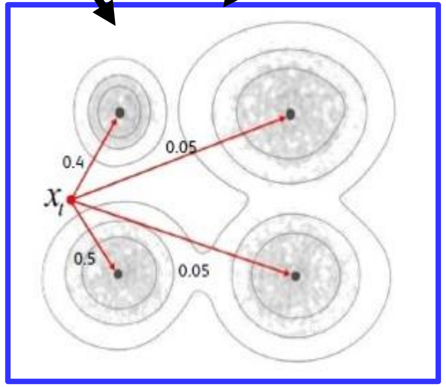


Universal GMM

$$\mathcal{G}_{\alpha_k}^X = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \left(\gamma_t(k) - w_k\right),$$

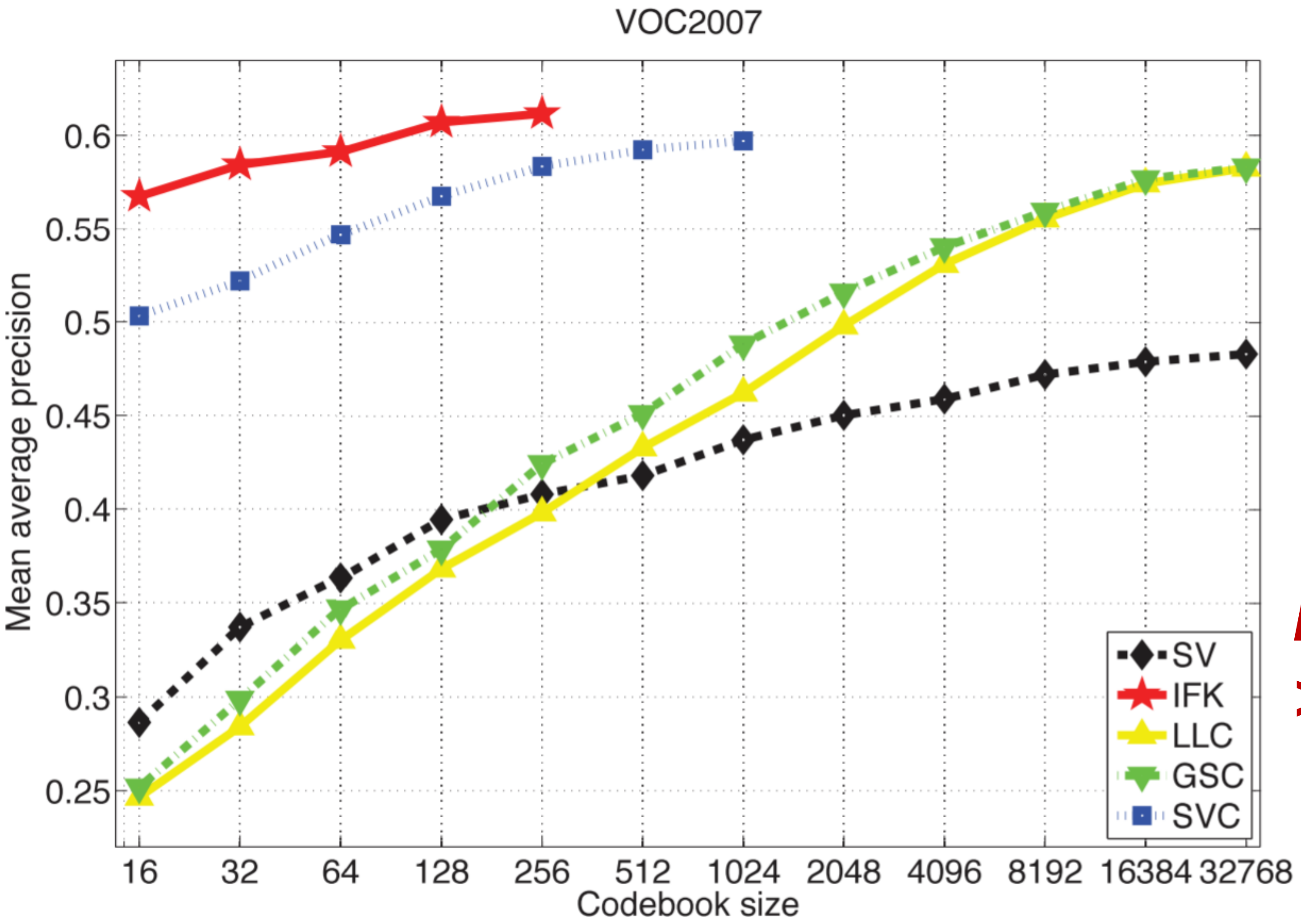
$$\mathcal{G}_{\mu_k}^X = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\frac{x_t - \mu_k}{\sigma_k}\right),$$

$$\mathcal{G}_{\sigma_k}^X = \frac{1}{\sqrt{w_k}} \sum_{t=1}^{T} \gamma_t(k) \frac{1}{\sqrt{2}} \left[\frac{(x_t - \mu_k)^2}{\sigma_k^2} - 1\right]$$



Higher-order Dictionary &
Higher-order Coding!

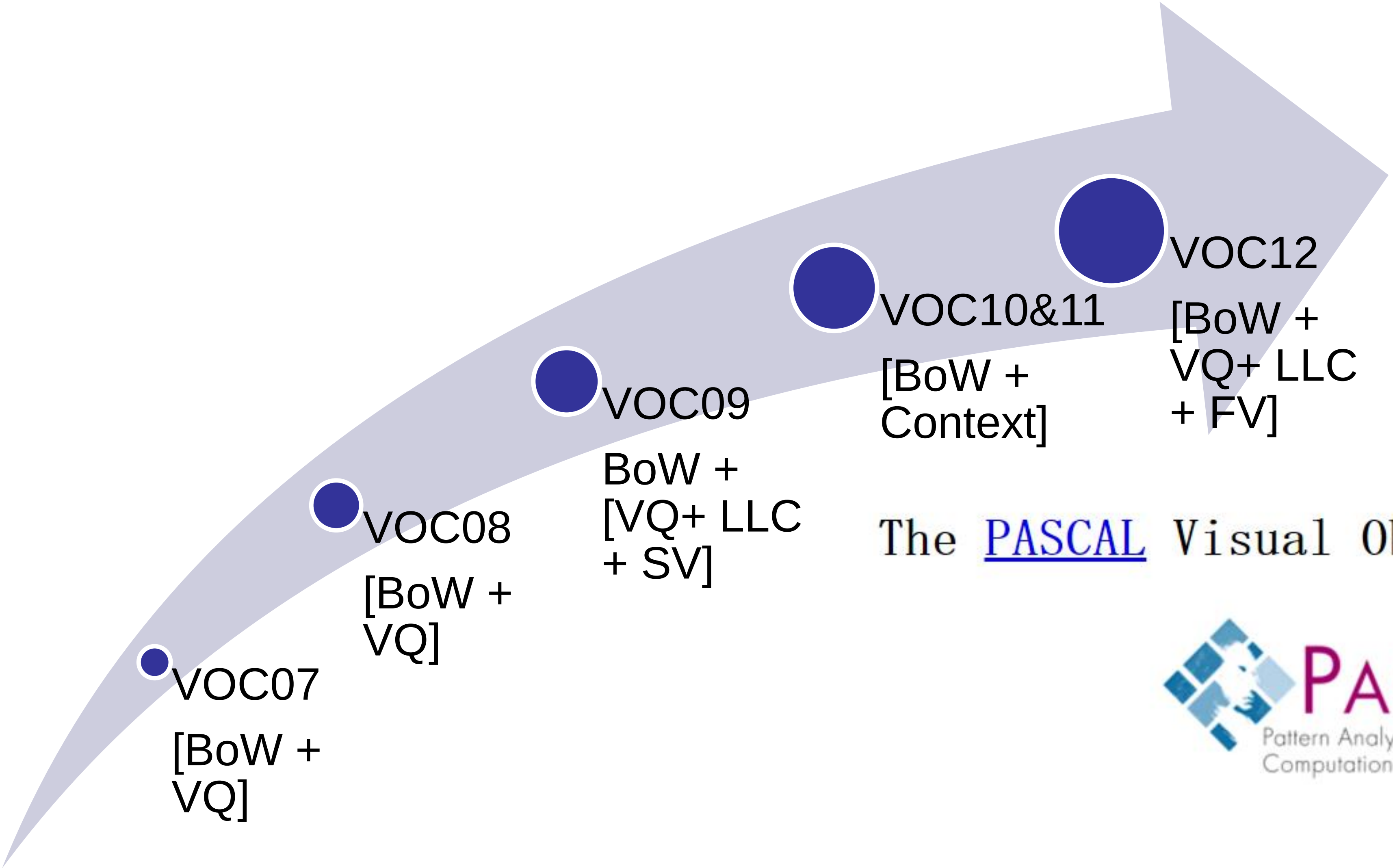
BoVW – Comparison



Yongzhen Huang, Zifeng Wu, Liang Wang, Tieniu Tan: Feature Coding in Image Classification: A Comprehensive Study. IEEE TPAMI. 36(3): 493-506 (2014)

Fisher Vector > Super Vector > Soft Coding > VQ!

BoVW – Comparison



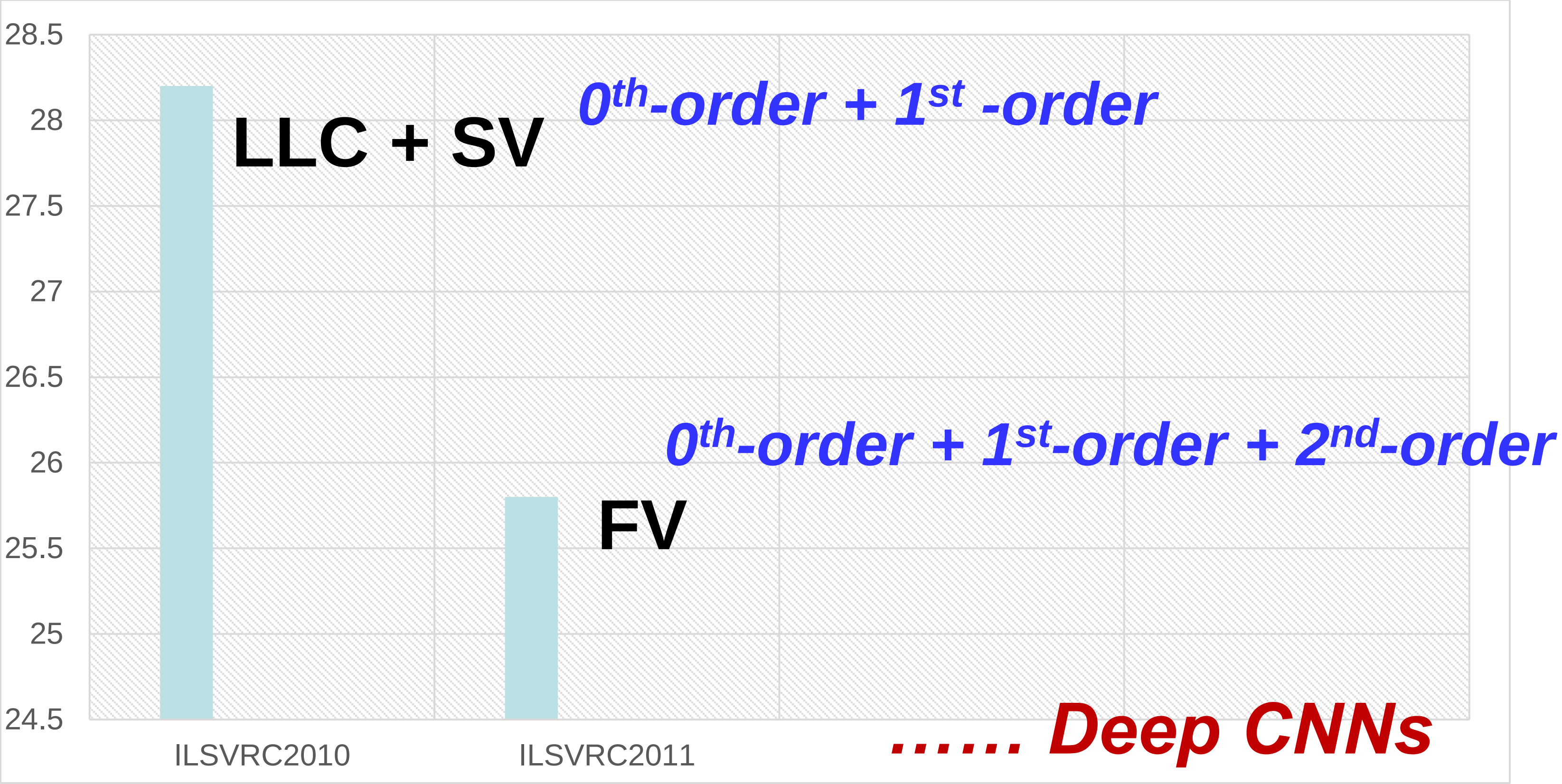
Fisher Vector
> Super Vector
> Soft Coding
> VQ!

The PASCAL Visual Object Classes Homepage

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

BoVW – Comparison

IMAGENET Large Scale Visual Recognition Challenge



BoVW – Comparison

FGComp'13 (Fine-Grained classification competition)

Team	Aircrafts	Birds	Cars	Dogs	Shoes	Overall
Ours: SA +SB	81.46	71.69	87.79	52.90	91.52	77.07
CafeNet*	78.85	73.01	79.58	57.53	90.12	75.82
Ours: SA	75.88	66.28	84.70	50.42	88.63	73.18
VisionMetric*	75.49	63.90	74.33	55.87	89.02	71.72
Symbiotic	75.85	69.06	81.03	44.89	87.33	71.63
Ours: SB	80.59	58.54	84.67	35.62	90.92	70.07
CognitiveVision*	67.42	72.79	64.39	60.56	84.83	70.00

Fisher Vector

AlexNet

BoVW – Higher-order VLAD

VLAD:

$$\mathbf{v}_k = N_k\left(\frac{1}{N_k}\sum_{j=1}^{N_k}\mathbf{x}_j - \mathbf{d}_k\right) = N_k(\mathbf{m}_k - \mathbf{d}_k)$$

2nd-order VLAD:

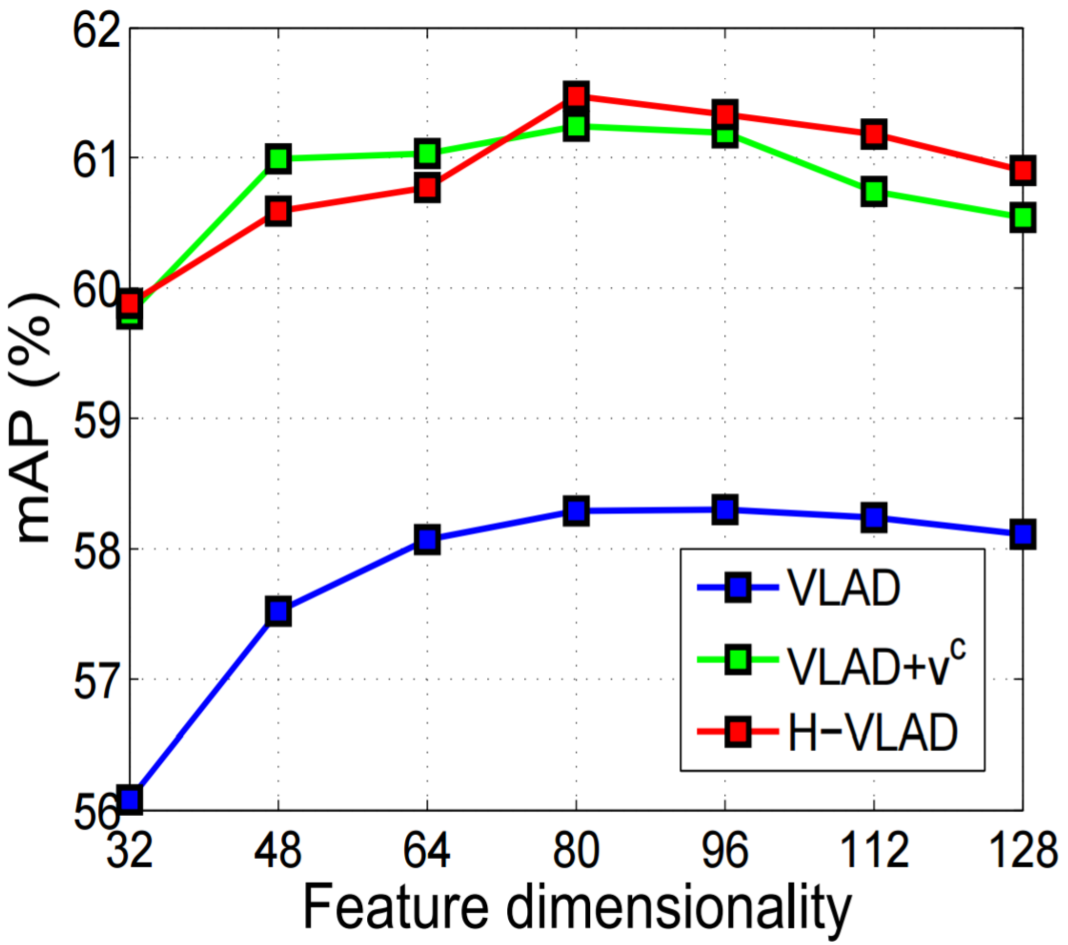
$$\mathbf{v}_k^c = \hat{\sigma}_k^2 - \sigma_k^2 = \frac{1}{N_k}\sum_{j=1}^{N_k}(\mathbf{x}_j - \mathbf{m}_k)^2 - \sigma_k^2,$$

3rd-order VLAD:

$$\mathbf{v}_k^s = \hat{\gamma}_k - \gamma_k = \frac{\frac{1}{N_k}\sum_{j=1}^{N_k}(\mathbf{x}_j - \mathbf{m}_k)^3}{\left(\frac{1}{N_k}\sum_{j=1}^{N_k}(\mathbf{x}_j - \mathbf{m}_k)^2\right)^{\frac{3}{2}}} - \gamma_k$$

Peng *et al.* Boosting VLAD with Supervised Dictionary Learning and High-Order Statistics. ECCV, 2014.

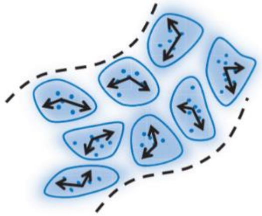
BoVW – Higher-order VLAD



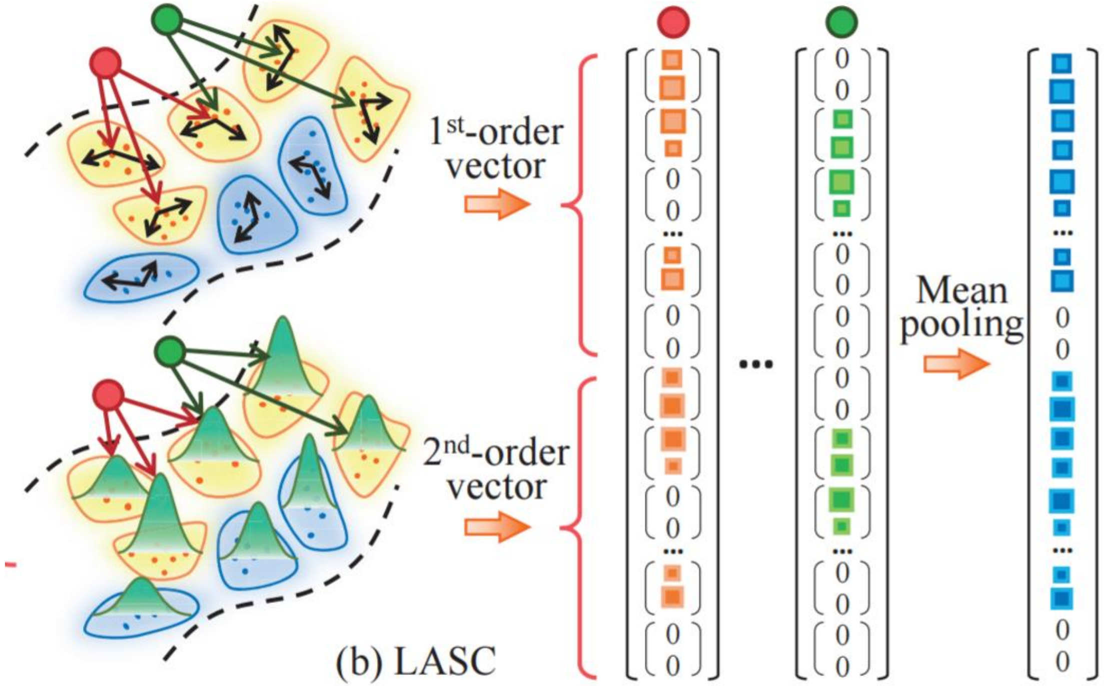
	HMDB51	UCF101
VLAD	55.5	84.8
H-VLAD	58.3	86.5
FV	58.5	86.7

Peng *et al.* Boosting VLAD with Supervised Dictionary Learning and High-Order Statistics. ECCV, 2014.

BoVW – Subspace Coding



Each atom is a Subspace.



(b) LASC

$$\mathbf{x} = \begin{bmatrix} \vdots \\ \mathbf{x}_i \\ \mathbf{x}_i^{:2} \\ \vdots \end{bmatrix}, \overbrace{\underbrace{\mathbf{x}_i = w_{\mathbf{y}}^i \begin{bmatrix} z_{i,1} \\ \vdots \\ z_{i,p} \end{bmatrix}}_{1^{\mathrm{st}}\text{-order vector}}, \underbrace{\mathbf{x}_i^{:2} = \frac{w_{\mathbf{y}}^i}{\sqrt{2}} \begin{bmatrix} \left(\frac{z_{i,1}}{\sigma_{i,1}}\right)^2 - 1 \\ \vdots \\ \left(\frac{z_{i,p}}{\sigma_{i,p}}\right)^2 - 1 \end{bmatrix}}_{2^{\mathrm{nd}}\text{-order vector}}}^{\text{LASC vector if } \mathcal{S}_i \in \mathcal{N}_k^S(\mathbf{y});\text{ otherwise } \mathbf{x}_i = \mathbf{x}_i^{:2} = \mathbf{0}}$$

$$\mathbf{x}_i = w_{\mathbf{y}}^i \mathbf{U}_i^T (\mathbf{y} - \boldsymbol{\mu}_i)$$

Subspace Dictionary
& Higher-order Coding!

Li *et al.* From Dictionary of Visual Words to Subspaces: Locality-constrained Affine Subspace Coding, CVPR, 2015.

BoVW – Subspace Coding

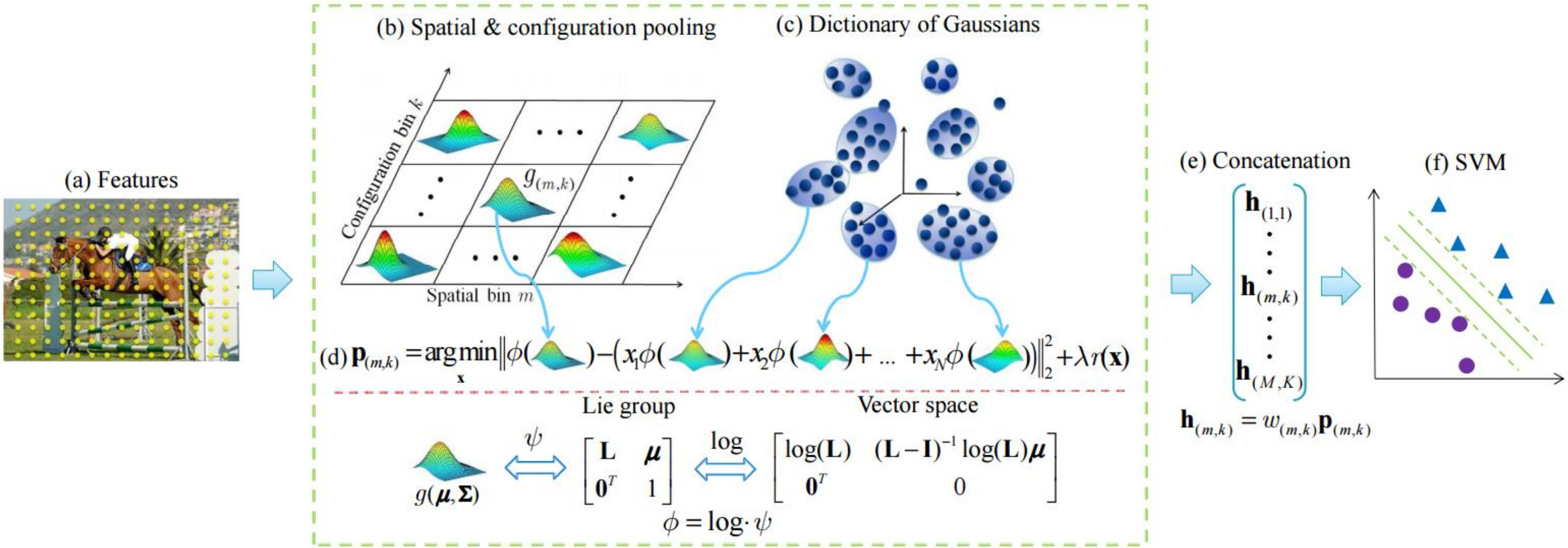
# samples	5	10	20	50
Xiao *et al.* [40]	14.5	20.9	28.1	38.0
LLC (4k) [16]	13.5	18.7	24.5	32.4
SV (128) [16]	16.4	21.9	28.4	36.6
FV (256) [31]	19.2 (0.4)	26.6 (0.4)	34.2 (0.3)	43.3 (0.2)
LASC (256)	**19.4 (0.4)**	**27.3 (0.3)**	**35.6 (0.1)**	**45.3 (0.4)**

Table 4. Comparison on SUN 397.

Li *et al.* From Dictionary of Visual Words to Subspaces: Locality-constrained Affine Subspace Coding, CVPR, 2015.

BoVW – Encoding Gaussians

Each atom is a Gaussian.



Encoding Gaussian over a Dictionary of Gaussians !

Li *et al*. High-order Local Pooling and Encoding Gaussians Over A Dictionary of Gaussians. IEEE TIP, 2017.

BoVW – Encoding Gaussians

# of train	5	10	20	50
Xiao et al. [27]	14.5	20.9	28.1	38.0
Kobayashi [12]	–	–	–	46.1 (0.1)
LASC [13]	19.4 (0.4)	27.3 (0.3)	35.6 (0.1)	45.3 (0.4))
FV (SIFT) [16]	19.2 (0.4)	26.6 (0.4)	34.2 (0.3)	43.3 (0.2)
FV (SIFT+LCS) [16]	21.1 (0.3)	29.1 (0.3)	37.4 (0.3)	47.2 (0.2)
HO-LP (SIFT)	21.9 (0.4)	29.9 (0.2)	37.6 (0.2)	47.1 (0.1)
HO-LP (SIFT+LCS)	25.7 (0.3)	34.6 (0.1)	42.9 (0.2)	51.4 (0.2)

Results on SUN 397

Li *et al*. High-order Local Pooling and Encoding Gaussians Over A Dictionary of Gaussians. IEEE TIP, 2017.

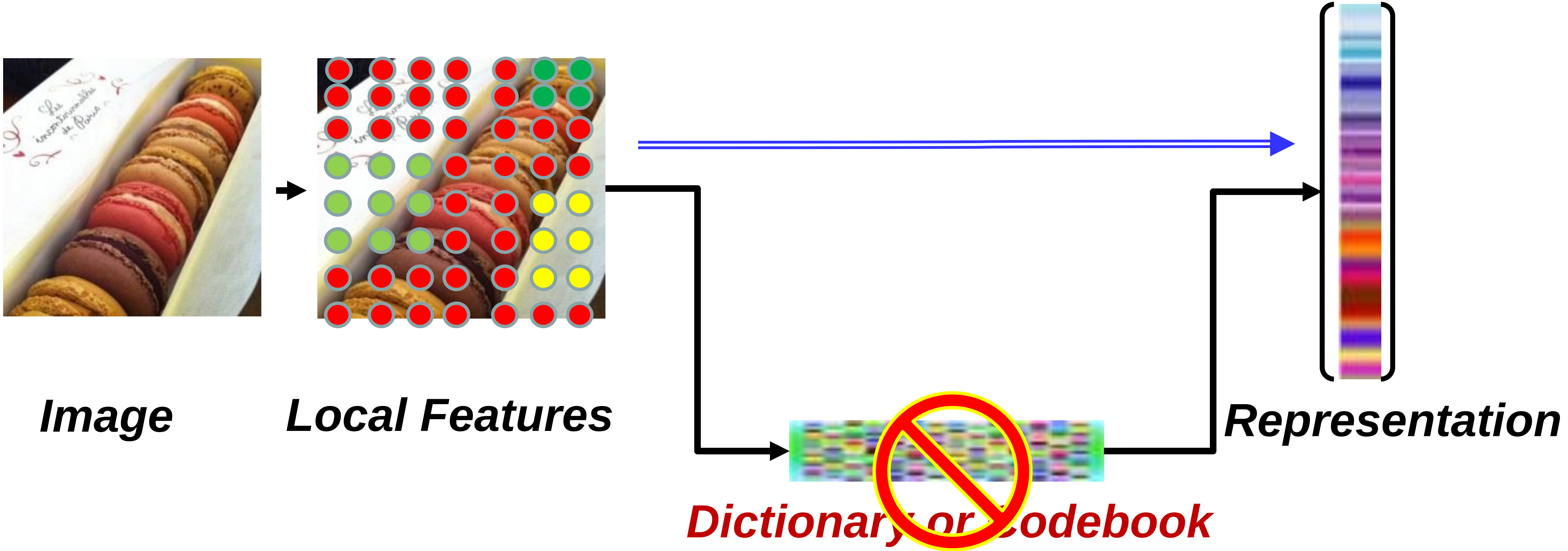
BoVW – Summary

- Bag-of-Visual-Words (BoVW) is a classical and popular model
- Perfromance: 1^{st} +2^{nd}-order coding > 1^{st}-order coding > 0^{th}-order coding
- **Higher-order Statistics** is **important** to Bag-of-Visual-Words (**BoVW**)

Context

1. Higher-order Statistics in Bag-of-Visual-Words (BoVW)
2. **Higher-order Statistics in Codebookless Model (CLM)**
3. Bag-of-Visual-Words *vs.* Codebookless Model
4. Higher-order Statistical Models Meet Deep Features

Codebookless Model (CLM)



CLM – Outline

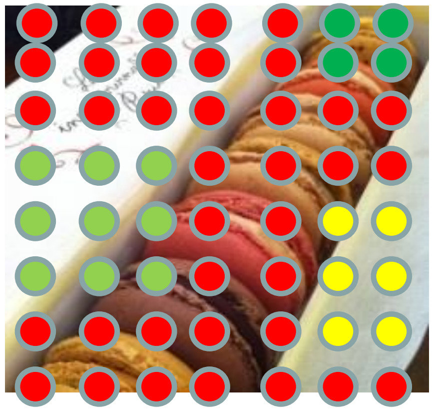
- Covariance Matrix (2^{nd} -order Statistics)
- Gaussian Model (1^{st} + 2^{nd} -order Statistics)
- Gaussian Mixture Model (1^{st} + 2^{nd} -order Statistics)
- 3-order Tensor Pooling (3^{rd}-order Statistics)

CLM – Covariance Matrix

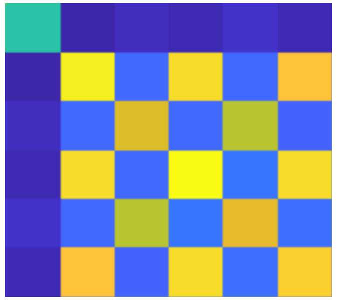
Application: Brain imaging [Arsigny et al 2005], Computer vision [Tuzel et al 2006], Machine learning [Kulis et al 2009], Radar signal processing [Barbaresco 2013].

Tuzel& Porikli& Meer [ECCV 2006, CVPR 2006, CVPR2008]: Modeling Image Regions with Covariance Matrices

Image or Patch



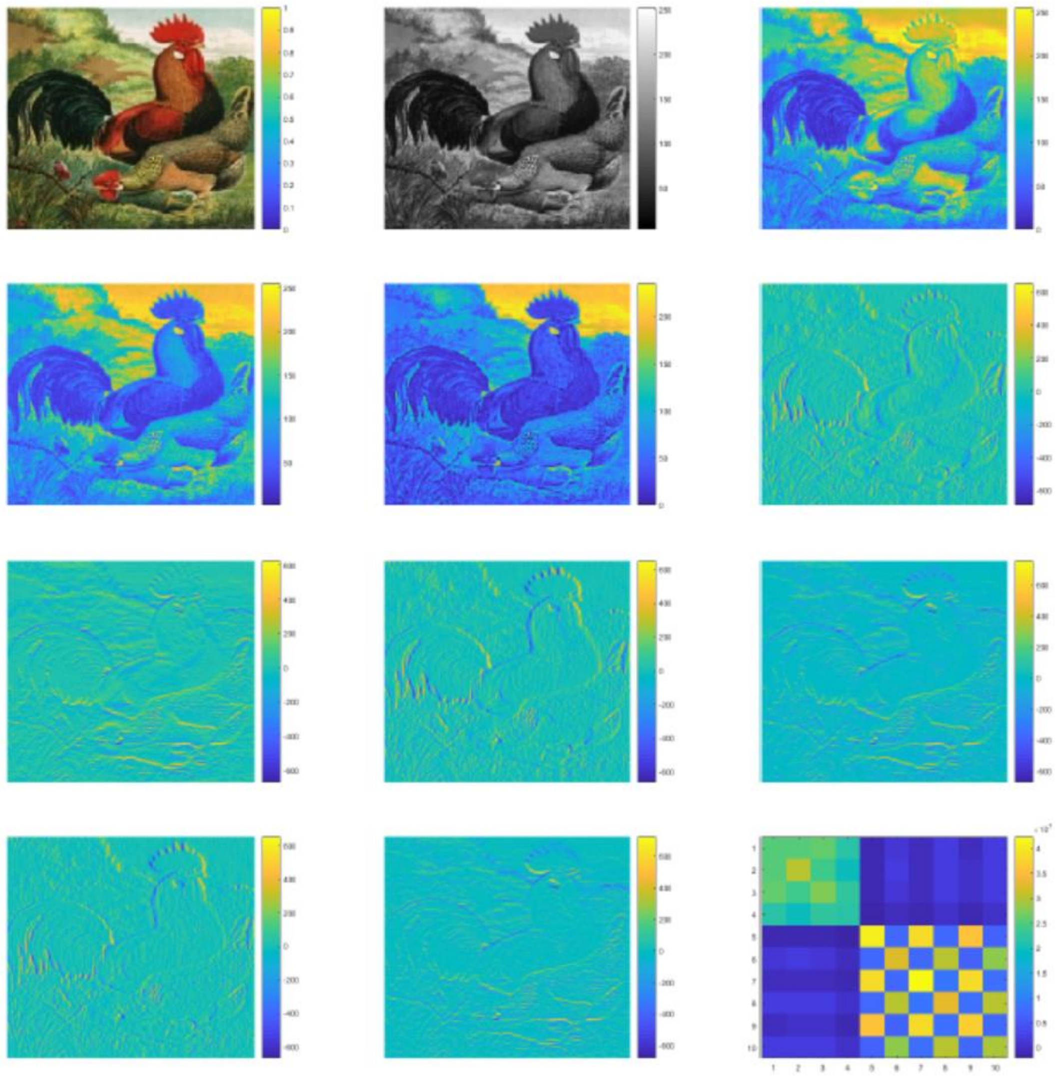
$$\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\right]$$



$$\boldsymbol{\Sigma} = \frac{1}{N}\mathbf{X}\mathbf{J}\mathbf{X}^T$$

$$\mathbf{J} = \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$$

CLM – Covariance Matrix

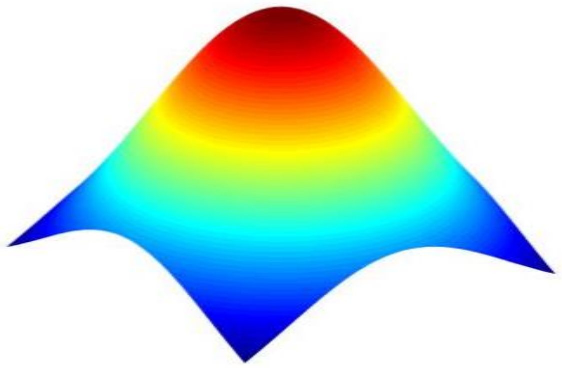


$$f(x,y)=\left[\begin{array}{l} I(x,y),R(x,y),G(x,y),B(x,y), \\ \left|\frac{\partial R}{\partial x}\right|,\left|\frac{\partial R}{\partial y}\right|,\left|\frac{\partial G}{\partial x}\right|,\left|\frac{\partial G}{\partial y}\right|,\left|\frac{\partial B}{\partial x}\right|,\left|\frac{\partial B}{\partial y}\right| \end{array}\right]$$

$$\mathbf{\Sigma}_{ij}=\left\langle \widehat{\mathbf{x}}^{i},\widehat{\mathbf{x}}^{j}\right\rangle$$

H`a Quang Minh. From Covariance Matrices to Covariance Operators: Data Representation from Finite to Infinite-Dimensional Settings. Tutorial ICCV, 2017.

CLM – Covariance Matrix



How to effectively and efficiently match two covariance matrices ?

CLM – Geometry of Covariance

- Euclidean space
  - Euclidean metric
- Riemannian manifold
  - Affine-invariant Riemannian metric
  - Log-Euclidean metric
- Convex cone
  - Bregman divergences

CLM – Geometry of Covariance

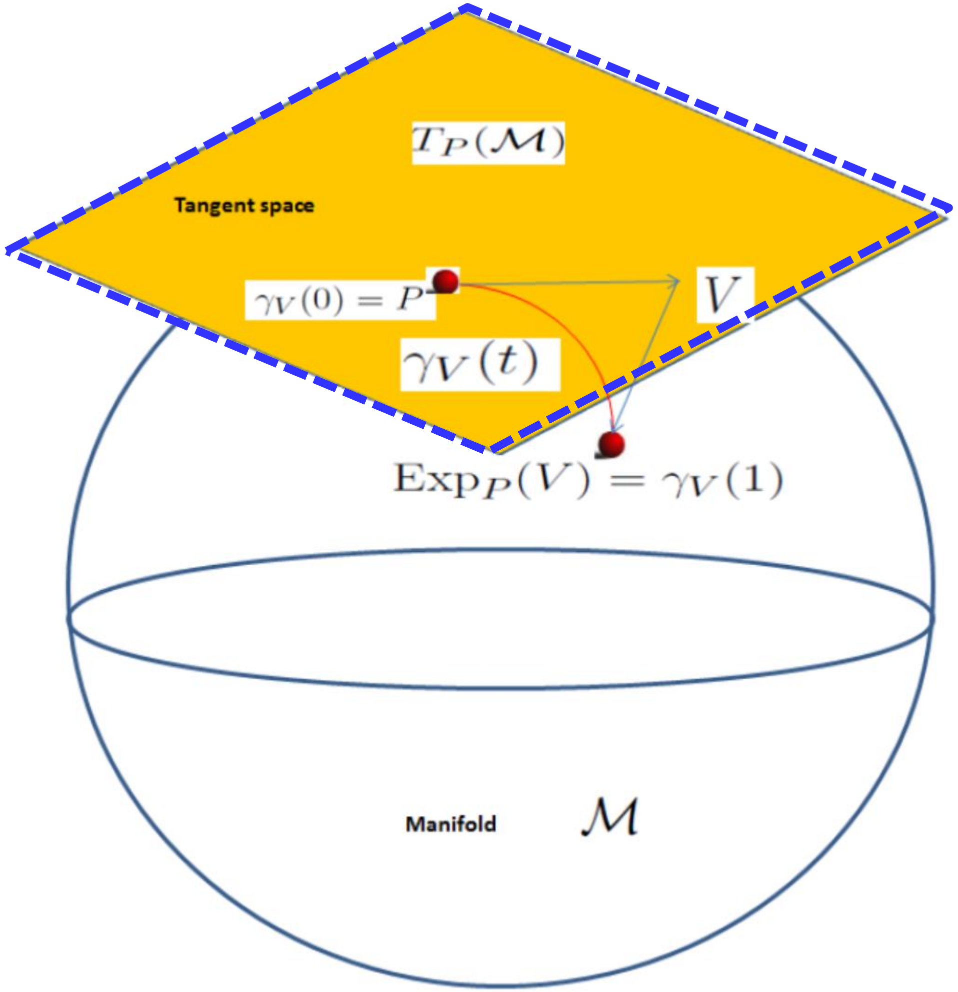
- Euclidean space

$$Sym^{+}(d) \subset Sym(d) \subset Mat(d)$$

$$d_E(A, B) = ||A - B||_F = ||\mathrm{vec}(A) - \mathrm{vec}(B)||$$

CLM – Geometry of Covariance

- Riemannian manifold



$$\boxed{L(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)} dt}$$

$$\boxed{\gamma_{AB}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}}$$

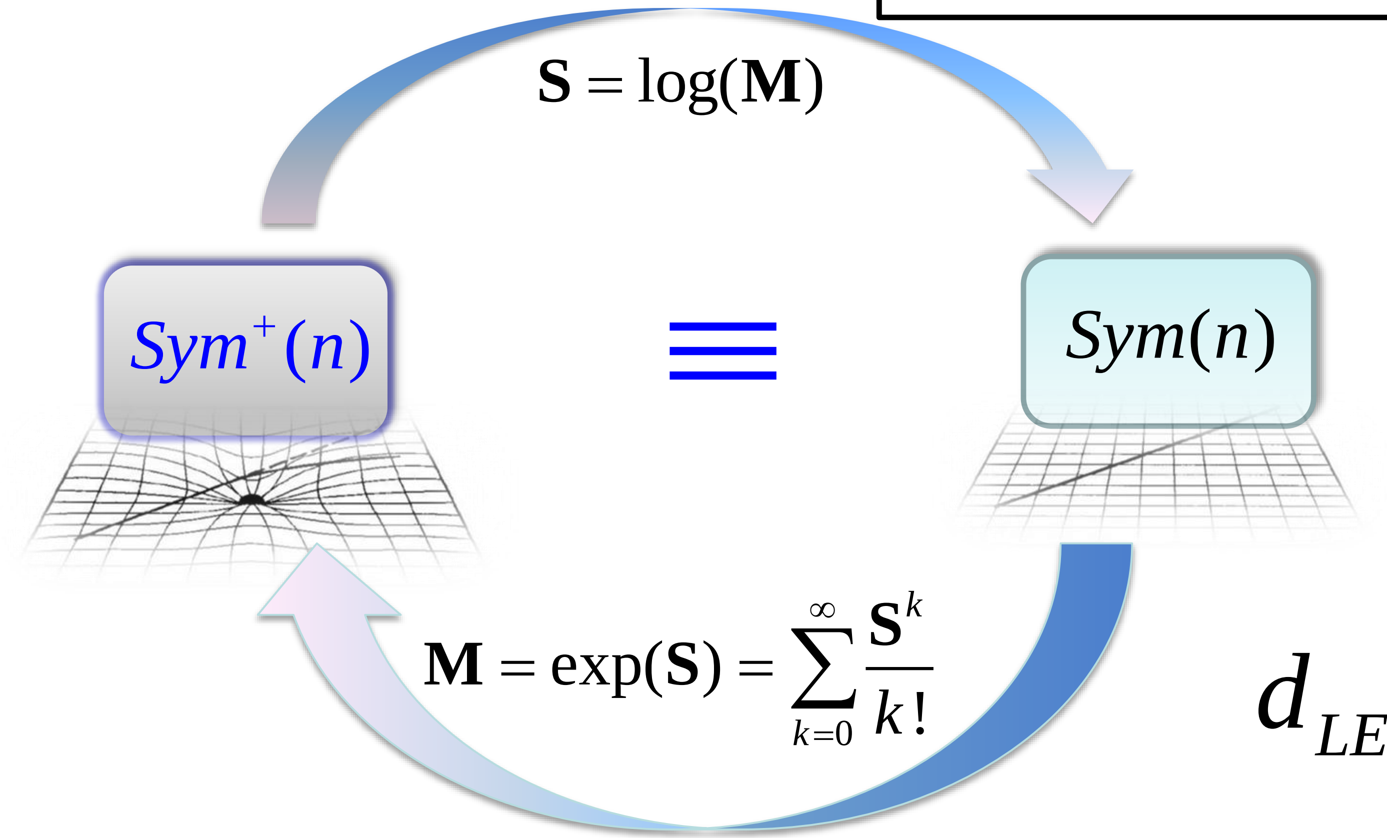
Affine-invariant Riemannian metric [Pennec et al 2006]:

$$d_{AIRM} = \left\| \log\left(\mathbf{A}^{-\frac{1}{2}} \mathbf{B} \mathbf{A}^{-\frac{1}{2}} \right) \right\|_F$$

CLM – Geometry of Covariance

- Riemannian manifold

$$\gamma_{AB}(t) = \exp[(1-t)\log(A) + t\log(B)]$$



Log-Euclidean Riemannian metric [Arsigny et al. 2007]:

$$d_{LERM} = \left\| \log(\mathbf{A}) - \log(\mathbf{B}) \right\|_F$$

CLM – Geometry of Covariance

- Convex cone

Ω = convex subset in $\mathbb{R}^n$

$\phi : \Omega \to \mathbb{R}$ = differentiable, strictly convex function

Bregman divergence on Ω (Bregman, 1967)

$$B(\mathbf{A},\mathbf{B}) = \phi(\mathbf{A}) - \phi(\mathbf{B}) - \langle \nabla\phi(\mathbf{A}), \mathbf{A}-\mathbf{B} \rangle$$

$$d_\phi^\alpha(\mathbf{A},\mathbf{B}) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\phi(\mathbf{A}) + \frac{1+\alpha}{2}\phi(\mathbf{B}) - \phi\left(\frac{1-\alpha}{2}\mathbf{A} + \frac{1+\alpha}{2}\mathbf{B}\right)\right],$$

$$-1 < \alpha < 1$$

CLM – Geometry of Covariance

$$\Omega = Sym^{++}, \quad \phi(\mathbf{A}) = -\log\det(\mathbf{A})$$ **[Linear Algebra and Its Applications, 2012]**

$$d_{\mathrm{logdet}}^{\alpha}(\mathbf{A},\mathbf{B}) = \frac{4}{1-\alpha^2}\log\frac{\det\left(\frac{1-\alpha}{2}\mathbf{A}+\frac{1+\alpha}{2}\mathbf{B}\right)}{\det(\mathbf{A})^{\frac{1-\alpha}{2}}\det(\mathbf{B})^{\frac{1+\alpha}{2}}}, -1<\alpha<1$$

$\alpha=0$ ***Symmetric Stein Divergence:***

$$d_{Stein} = 4\left[\log\det\left(\frac{\mathbf{A}+\mathbf{B}}{2}\right) - \frac{1}{2}\log\det(\mathbf{AB})\right]$$

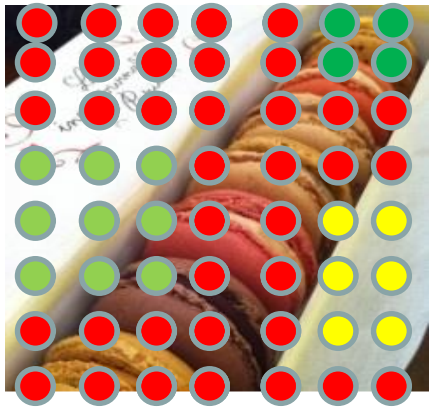
CLM – Geometry of Covariance

	Euclidean	ARIM	LERM	LogDet
Geodesic Distance	Yes	Yes	Yes	No
Invariance	No	Affine	Similarity	Affine
Inner Product Distance	Yes	No	Yes	No
Decoupled	Yes	No	Yes	No
Computational Cost	Fastest	Slow	Fast	Fast

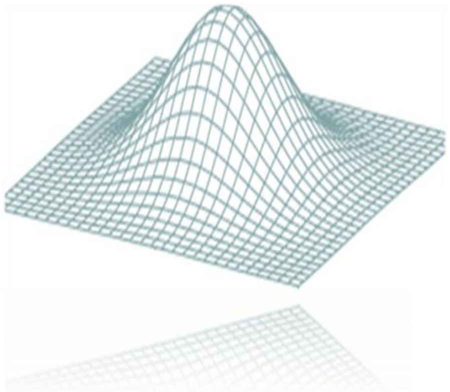
CLM – Gaussian Model

Nakayama et al. [CVPR10]
Wang et al. [PR16]
Wang et al. [CVPR16]

Image or Patch



$$\mathbf{X}=\left[\mathbf{x}_1,\mathbf{x}_2,\ldots\mathbf{x}_N\right]$$



$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$$\boldsymbol{\mu}=\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k,\quad \boldsymbol{\Sigma}=\frac{1}{N-1}\mathbf{X}\mathbf{J}\mathbf{X}^T.$$

CLM – Matching Gaussian Models

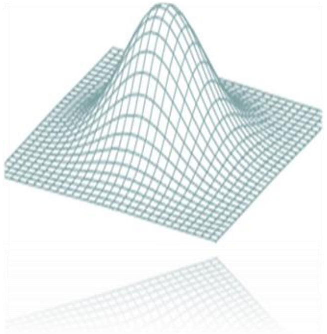
- How to Match Gaussian Models ?

  - Information geometry

  - Embedded Riemannian manifold

  - Lie group theory

CLM – Geometry of Gaussian



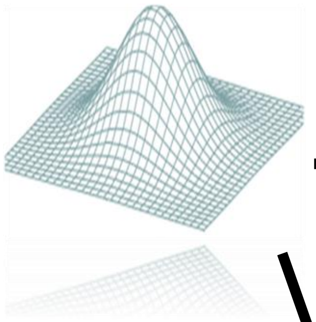
$$\boldsymbol{\eta} = (\hat{\mu}_1, ..., \hat{\mu}_d, \hat{\Sigma}_{11} + \hat{\mu}_1^2, ..., \hat{\Sigma}_{1d} + \hat{\mu}_1\hat{\mu}_d, \hat{\Sigma}_{22} + \hat{\mu}_2^2, ..., \hat{\Sigma}_{dd} + \hat{\mu}_d^2)^T.$$

Euclidean Kernel:	$\boldsymbol{\eta}(P)^T\boldsymbol{\eta}(Q)$
Center Tangent Kernel:	$\boldsymbol{\eta}(P)^T G^{\eta}(\boldsymbol{\eta}_c)\boldsymbol{\eta}(Q)$
KL-divergence:	$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \mathrm{tr}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) - 2n$

[1] H. Nakayama et al, Global Gaussian approach for scene categorization using information geometry. CVPR, 2010.
[2] S. ichi Amari and H. Nagaoka, Methods of Information Geometry. London, U.K.: Oxford Univ. Press, 2000.

CLM – Geometry of Gaussian

$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$



$$\mathbf{B} = \begin{bmatrix} \widetilde{\mathbf{L}} & \mu \\ \mathbf{0}^T & 1 \end{bmatrix}$$

Affine Group [Gong et al. CVPR09]

$$\mathbf{B} = \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix}$$

Siegel Group [Calvo et al. JMV 1990]

$$\mathbf{B} = |\Sigma|^{-\frac{2}{n+1}} \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix}$$

Riemannian Symmetric Group [Lovric et al. JMV 2000]

$$\left\| \log\left(\mathbf{B}_1^{-1} \mathbf{B}_2 \right) \right\|_F$$

CLM – Geometry of Gaussian

Definition 1. *Let $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \in N(n), i = 1, 2$, be two arbitrary Gaussians and $\boldsymbol{\Sigma}_i = \mathbf{L}_i^{-T}\mathbf{L}_i^{-1}$, where $\mathbf{L}_i$ is the Cholesky factor of $\boldsymbol{\Sigma}_i^{-1}$. We define an operation $\star$ between two Gaussians as*

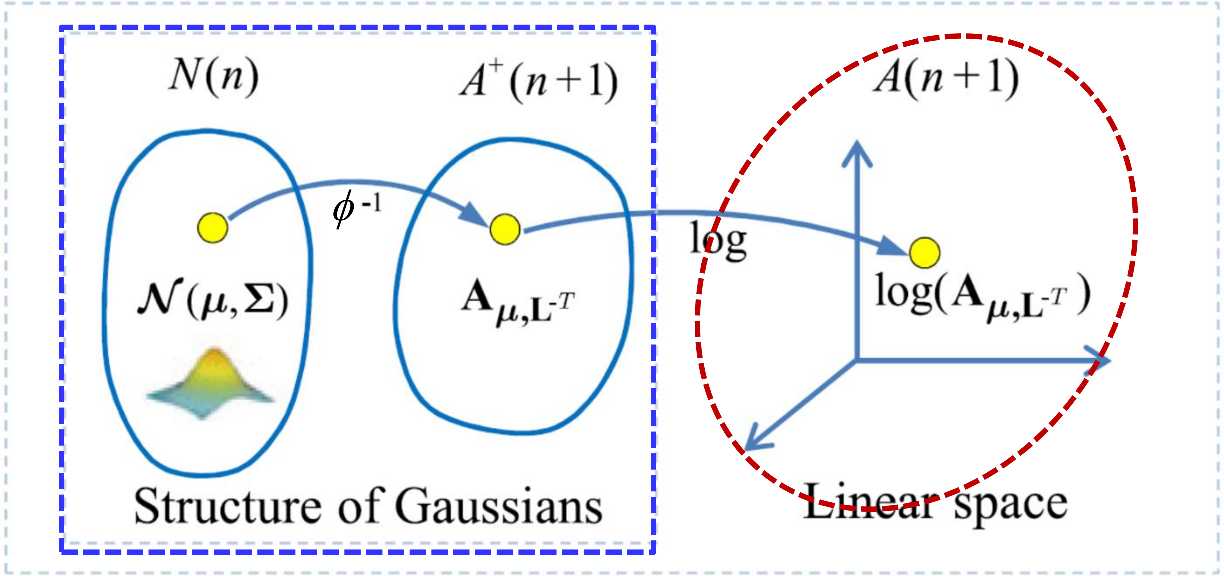
$$\begin{aligned}\star :& N(n) \times N(n) \to N(n) \\ & \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \star \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \\ & \quad = \mathcal{N}(\mathbf{L}_1^{-T}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1, (\mathbf{L}_1\mathbf{L}_2)^{-T}(\mathbf{L}_1\mathbf{L}_2)^{-1}).\end{aligned} \tag{3}$$

Theorem 1. *$N(n)$ is a Lie group under multiplication operation $\star$ as defined in (3).*

Peihua Li, Qilong Wang *et al.* Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification. TPAMI, 2017.

CLM – Geometry of Gaussian

Space of Gaussians is equipped with a Lie group structure.



$$\Sigma^{-1} = \mathbf{L}\mathbf{L}^{T}$$

$$\mathcal{N}(\mu,\Sigma) \xrightarrow{\phi^{-1}} \mathbf{A}_{\mu,\mathbf{L}^{-T}} = \begin{bmatrix} \mathbf{L}^{-T} & \mu \\ \mathbf{0}^{T} & 1 \end{bmatrix}$$

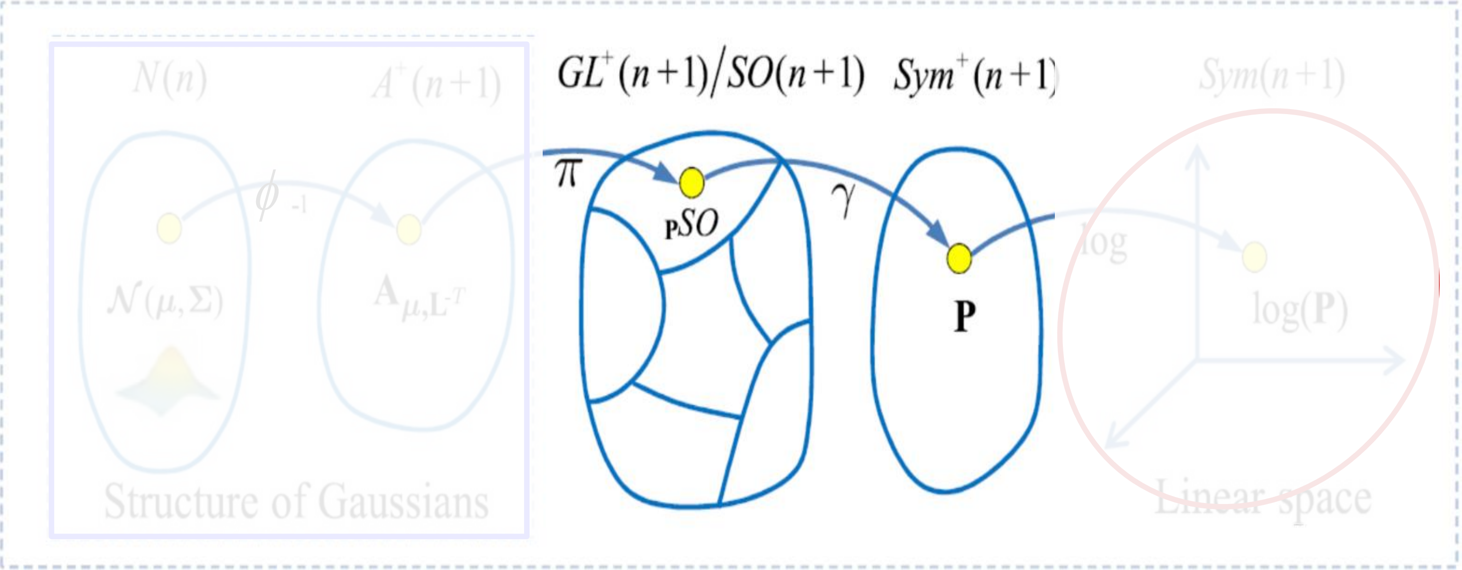
Lie group as well

$$\log\left(\mathbf{A}_{\mu,\mathbf{L}^{-T}}\right) = \log\left(\begin{bmatrix} \mathbf{L}^{-T} & \mu \\ \mathbf{0}^{T} & 1 \end{bmatrix}\right)$$

LERM on $A^{+}(n+1)$

Peihua Li, Qilong Wang *et al.* Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification. TPAMI, 2017.

CLM – Geometry of Gaussian



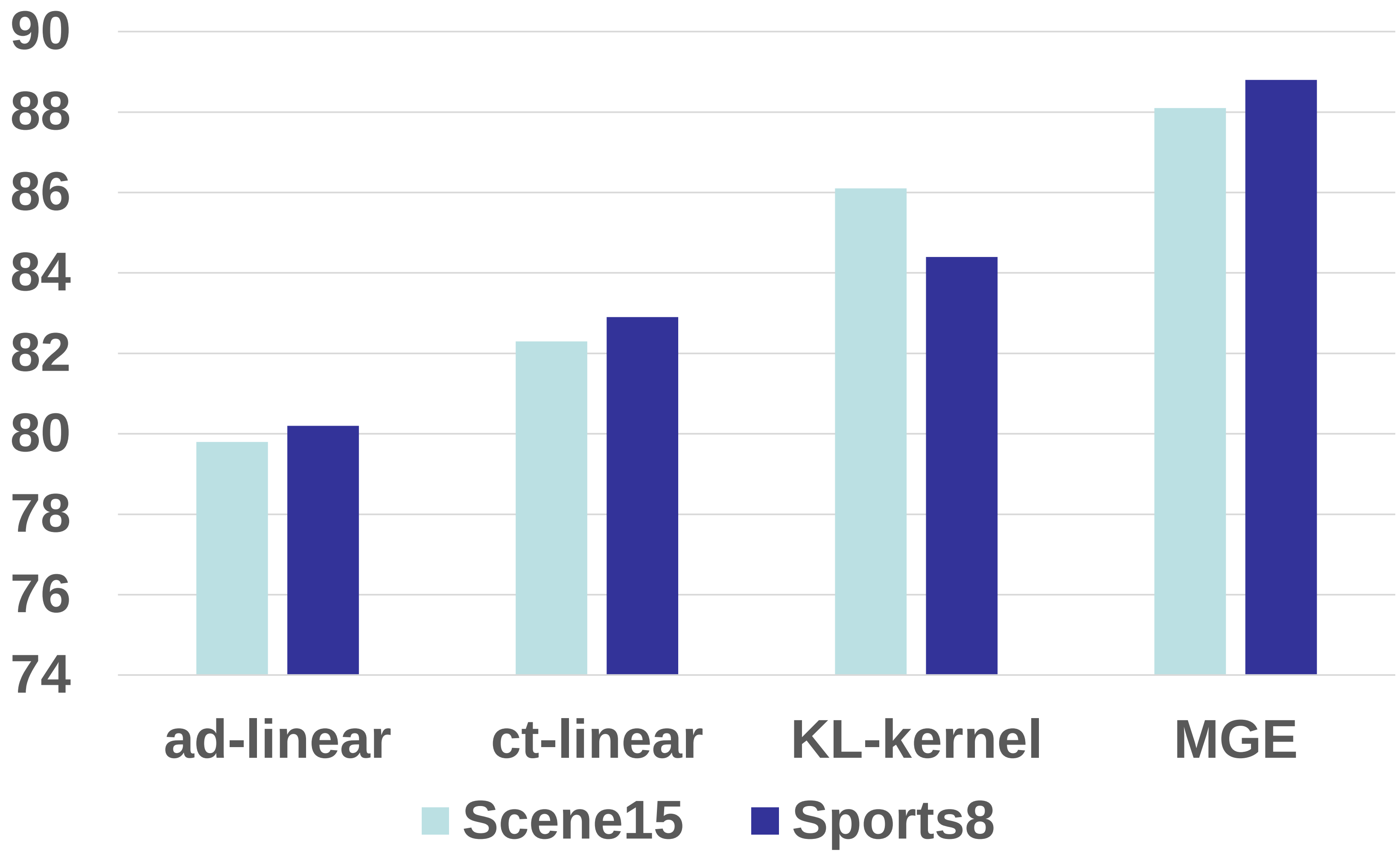
$$\mathcal{N}(\mu,\Sigma) \xrightarrow{\phi^{-1}} \mathbf{A}_{\mu,\mathbf{L}^{-T}} = \begin{bmatrix} \mathbf{L}^{-T} & \mu \\ \mathbf{0}^{T} & 1 \end{bmatrix}$$

SPD manifold

$$\begin{bmatrix} \mathbf{L}^{-T} & \mu \\ \mathbf{0}^{T} & 1 \end{bmatrix} \xrightarrow{\pi} \mathbf{PO} \xrightarrow{\gamma} \mathbf{P} = \begin{bmatrix} \Sigma + \mu\mu^{T} & \mu \\ \mu^{T} & 1 \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} \mathbf{L}^{-T} & \mu \\ \mathbf{0}^{T} & 1 \end{bmatrix} \xrightarrow{\pi} \mathbf{OP} \xrightarrow{\gamma} \mathbf{P} = \begin{bmatrix} \mathbf{L}^{-1}\mathbf{L}^{-T} & \mathbf{L}^{-1}\mu \\ \mu^{T}\mathbf{L}^{-T} & \mu^{T}\mu + 1 \end{bmatrix}^{\frac{1}{2}}$$

Peihua Li, Qilong Wang *et al.* Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification. TPAMI, 2017.

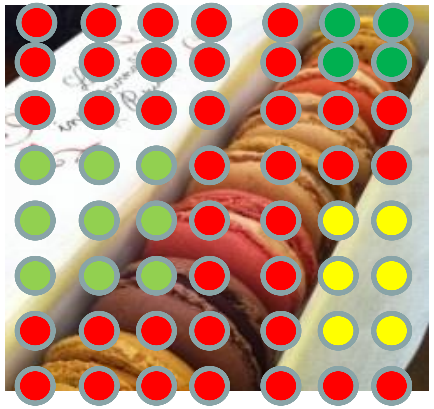
CLM – Geometry of Gaussian



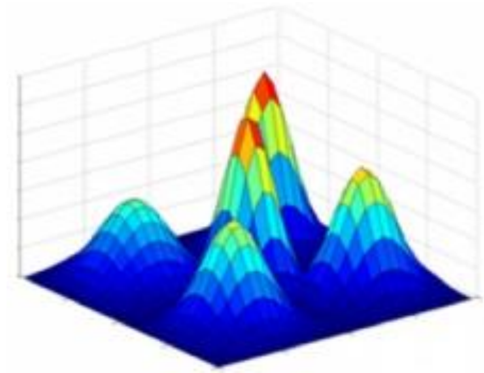
Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

CLM – Gaussian Mixture Model (GMM)

Image or Patch



$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_N]$

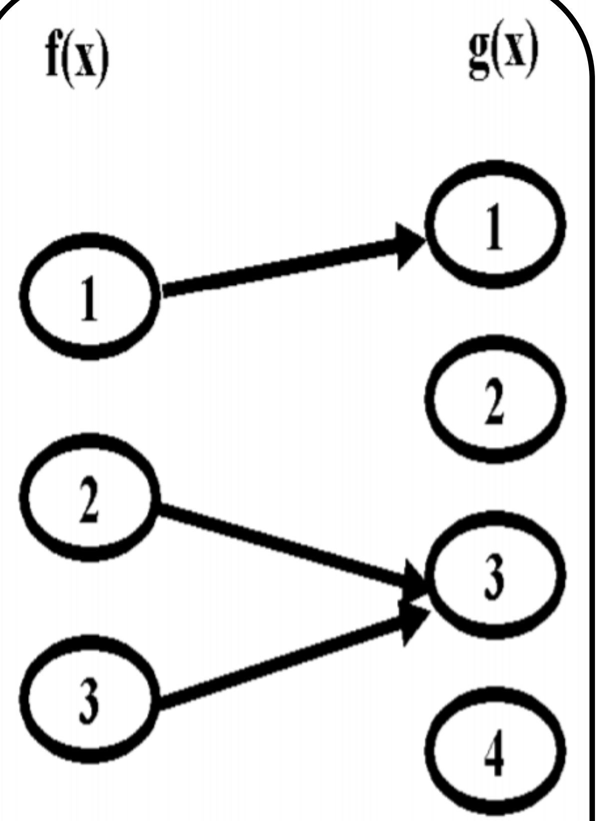


$$G(\mathbf{f}) = \sum_{i=1}^{n} w_i \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

[Goldberger et al. ICCV 03]
[Beecks et al. ICCV 11]
[Li et al. ICCV 13]

Measures for GMMs ?

CLM – Gaussian Mixture Model (GMM)



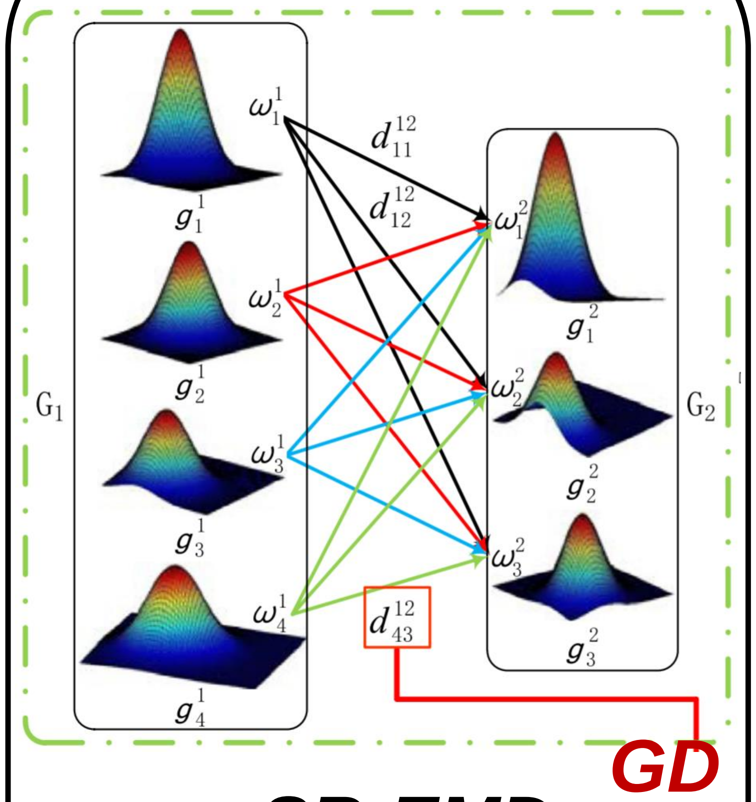
Match-KL
[ICCV 03]

$$\mathrm{GQFD}_{f_s}(g_a, g_b)$$

$$\sqrt{(w^a| - w^b) \cdot A_{f_s} \cdot (w^a| - w^b)^T}$$

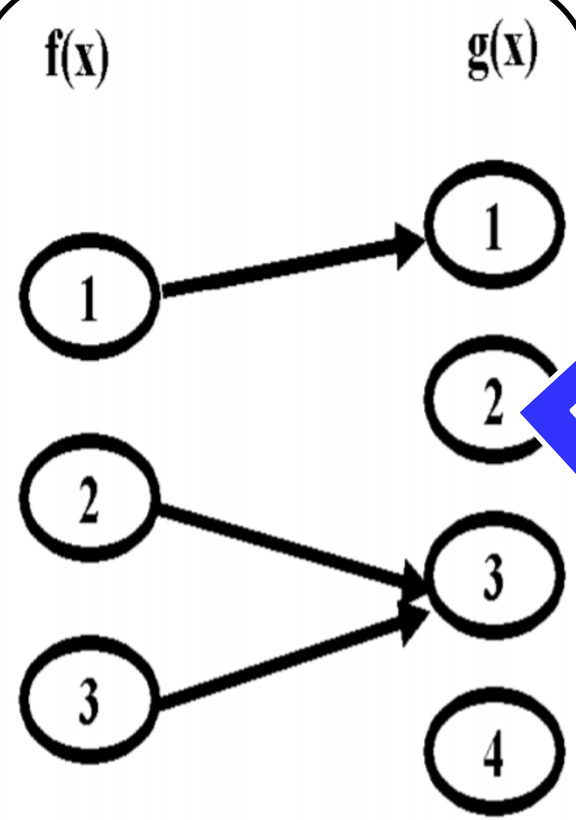
$$\approx \sum_{a,b} w_{ab} \left\| g_a - g_b \right\|$$

GQFD
[ICCV 11]



SR-EMD
[ICCV 13]

CLM – Gaussian Mixture Model (GMM)



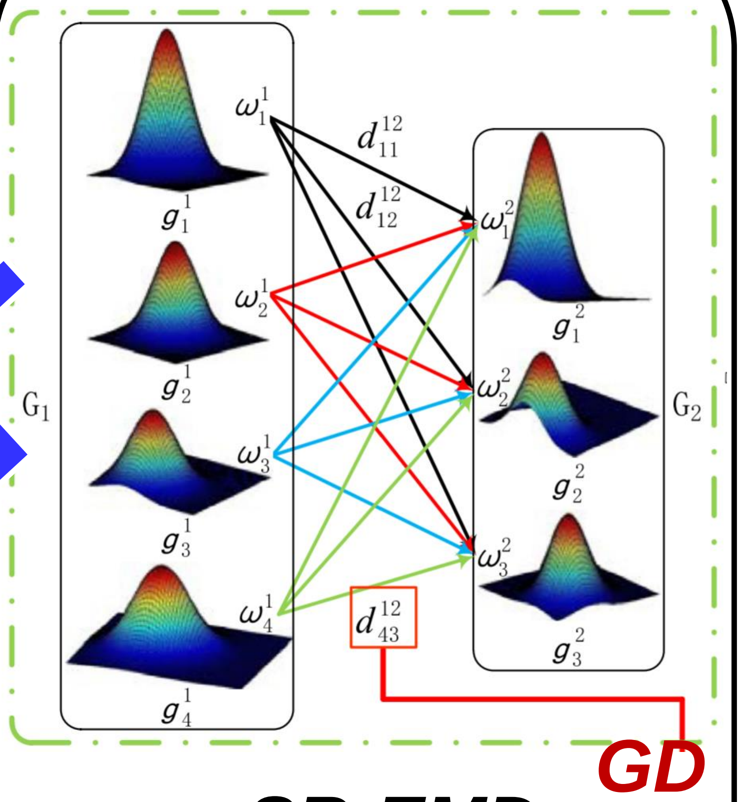
Match-KL
[ICCV 03]

$$\mathrm{GQFD}_{f_s}(g_a, g_b)$$

$$\sqrt{(w^a| - w^b) \cdot A_{f_s} \cdot (w^a| - w^b)^T}$$

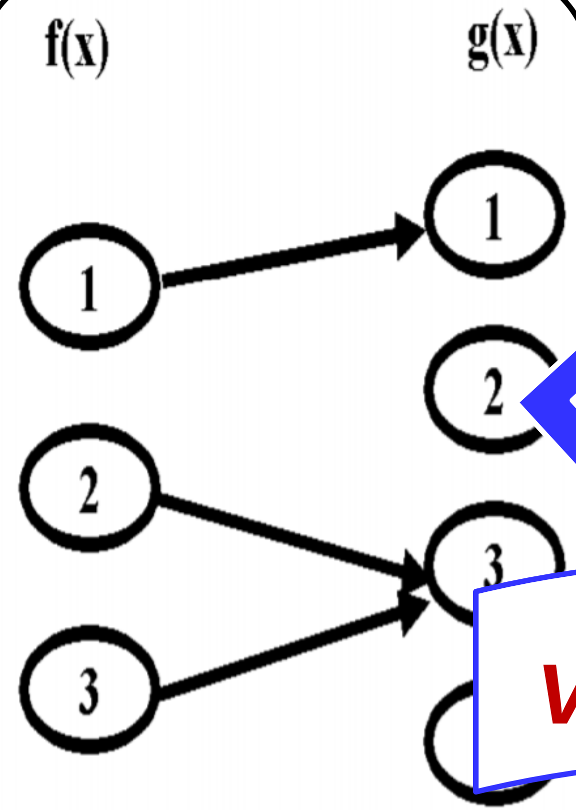
$$\approx \sum_{a,b} w_{ab} \left\| g_a - g_b \right\|$$

GQFD
[ICCV 11]



SR-EMD
[ICCV 13]

CLM – Gaussian Mixture Model (GMM)

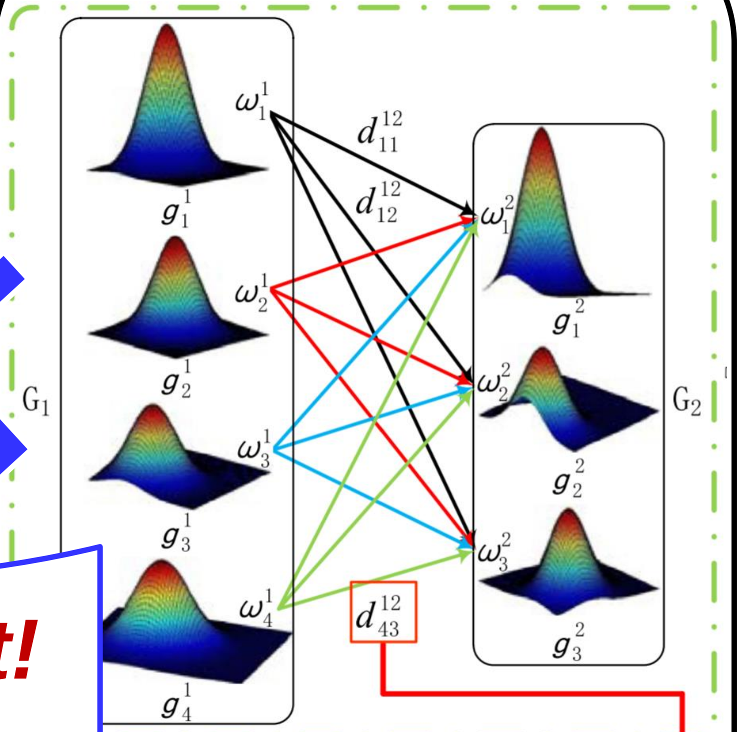


Match-KL
[ICCV 03]

$$\mathrm{GQFD}_{f_s}(g_a, g_b)$$

$$\sqrt{(w^a| - w^b) \cdot A_{f_s} \cdot (w^a| - w^b)^T}$$

GQFD
[ICCV 11]



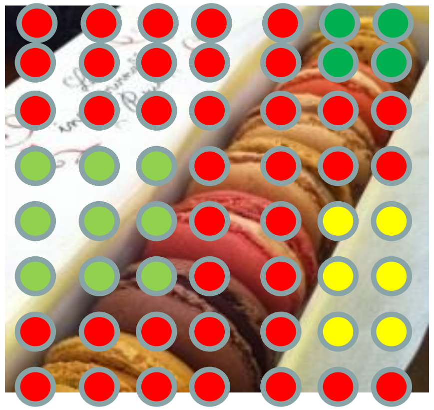
GD

SR-EMD
[ICCV 13]

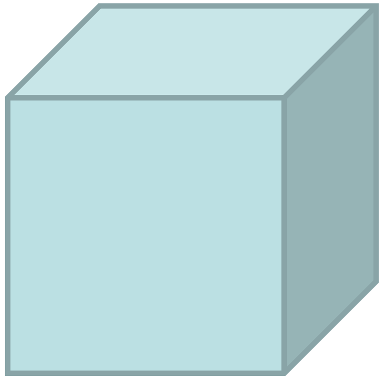
Very High Computational Cost!

CLM – 3-order Tensor Pooling

Image or Patch



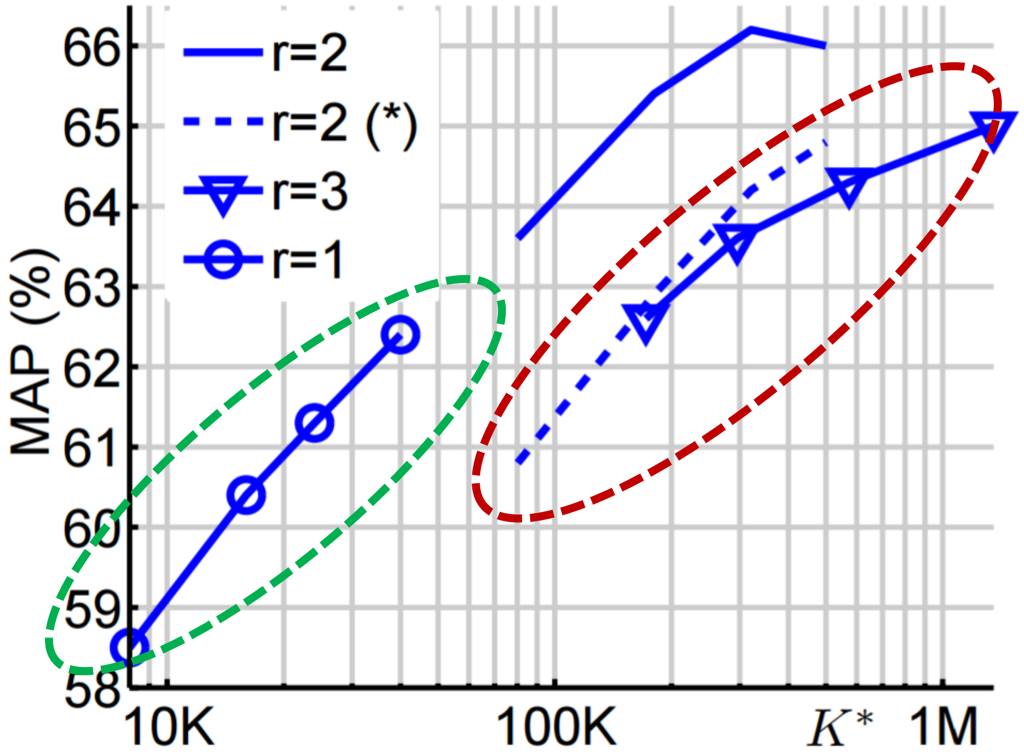
$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N]$



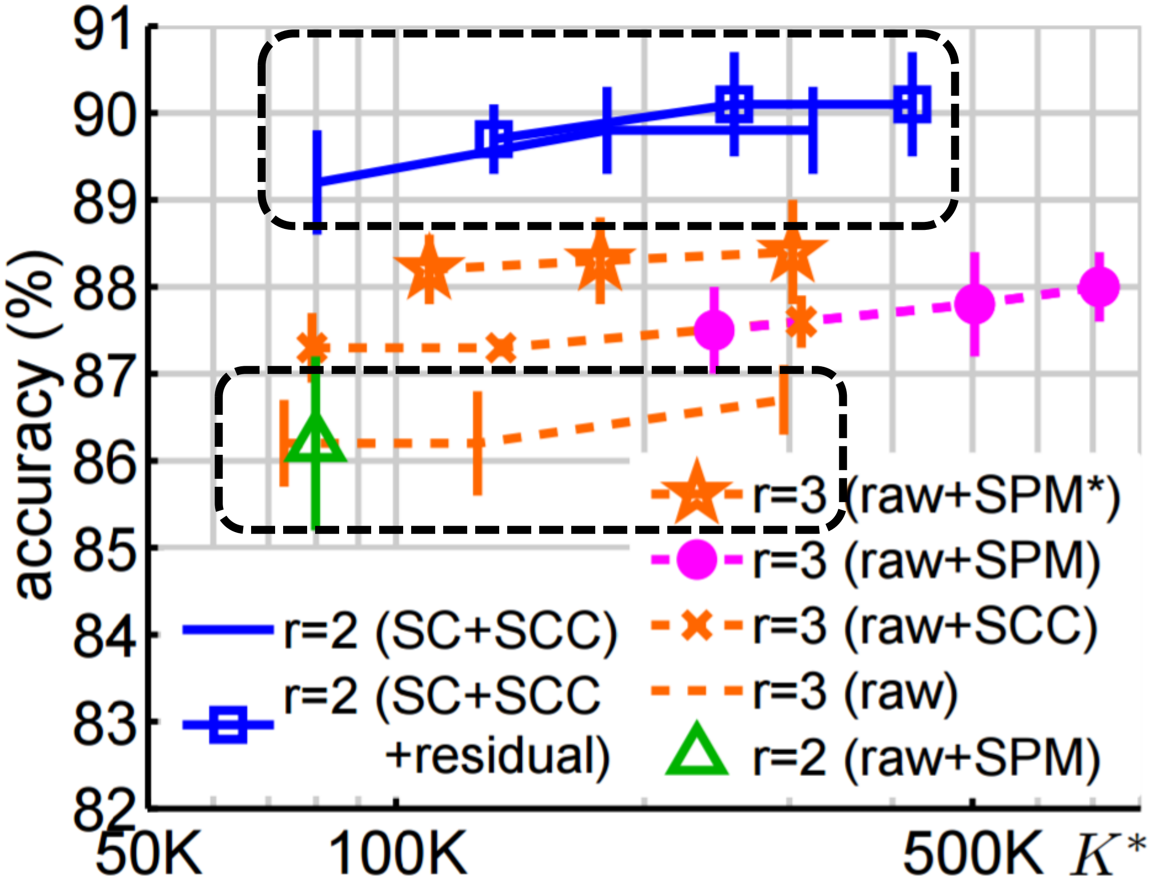
$\mathbf{X} \otimes \mathbf{X} \otimes \mathbf{X}$

Higher-order Occurrence Pooling on Mid- and Low-level Features: Visual Concept Detection. TPAMI, 2018.

CLM – Comparison



(a) VOC07, $r = 1, 2, 3$



Higher-order Occurrence Pooling on Mid- and Low-level Features: Visual Concept Detection. TPAMI, 2018.

CLM – Summary

- Higher-order CLM has special (non-Euclidean) geometry structure.
- Higher-order CLM leads higher dimensional representations, and appropriate higher-order statistics bring better performance.
- Compared with BoVW, CLM attracts much less attentions.

Context

1. Higher-order Statistics in Bag-of-Visual-Words (BoVW)
2. Higher-order Statistics in Codebookless Model (CLM)
3. **Bag-of-Visual-Words *VS.* Codebookless Model**
4. Higher-order Statistical Models Meet Deep Features

BoVW *VS.* CLM

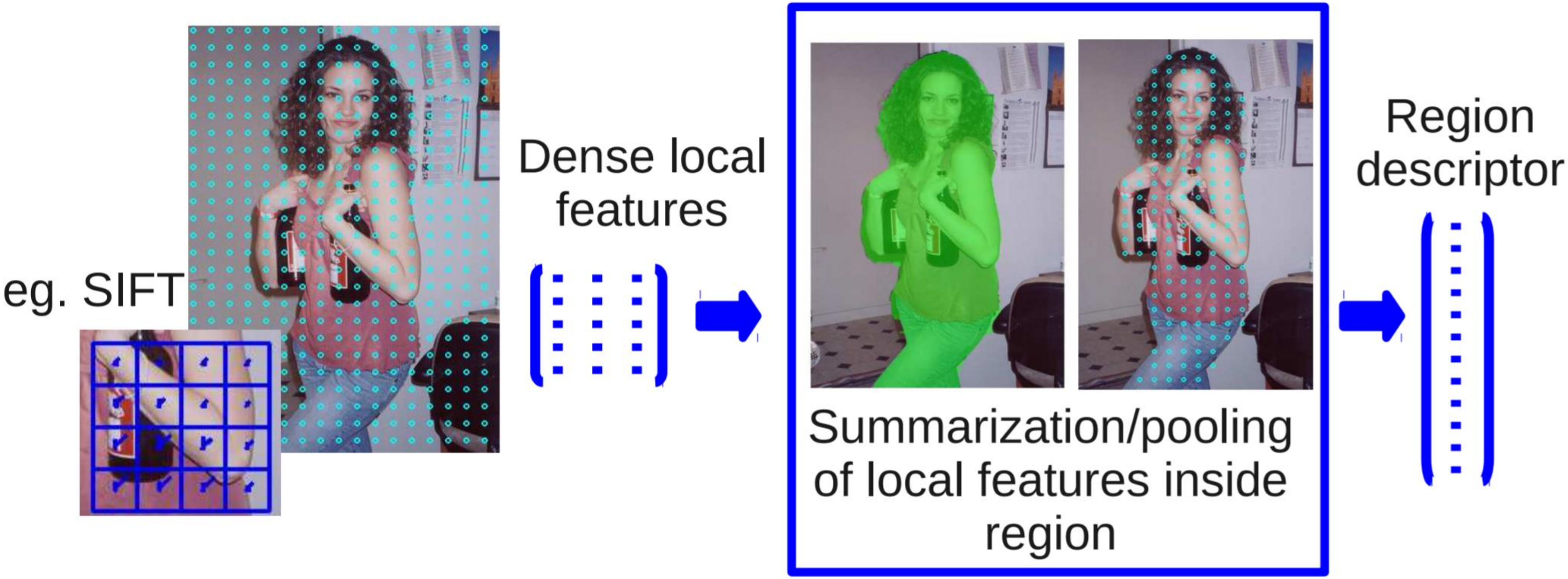
- Limitations of BoVW
  - The codebook brings quantization error. [Boiman et al. CVPR08]
  - Training & coding large-size codebook is time-consuming . An real universal codebook is unavailable.
  - Assumption of channel intendent in high-order statistics.
- Limitations of CLM
  - Measuring CLM is usually high computational cost.
  - CLM seems inferior to BoVW for computer vision tasks.

BoVW *VS.* CLM

- Free-form Region Modeling
  - J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

- Whole Image Modeling
  - Qilong Wang, Peihua Li, Wangmeng Zuo, Lei Zhang. Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Free-form Region Modeling



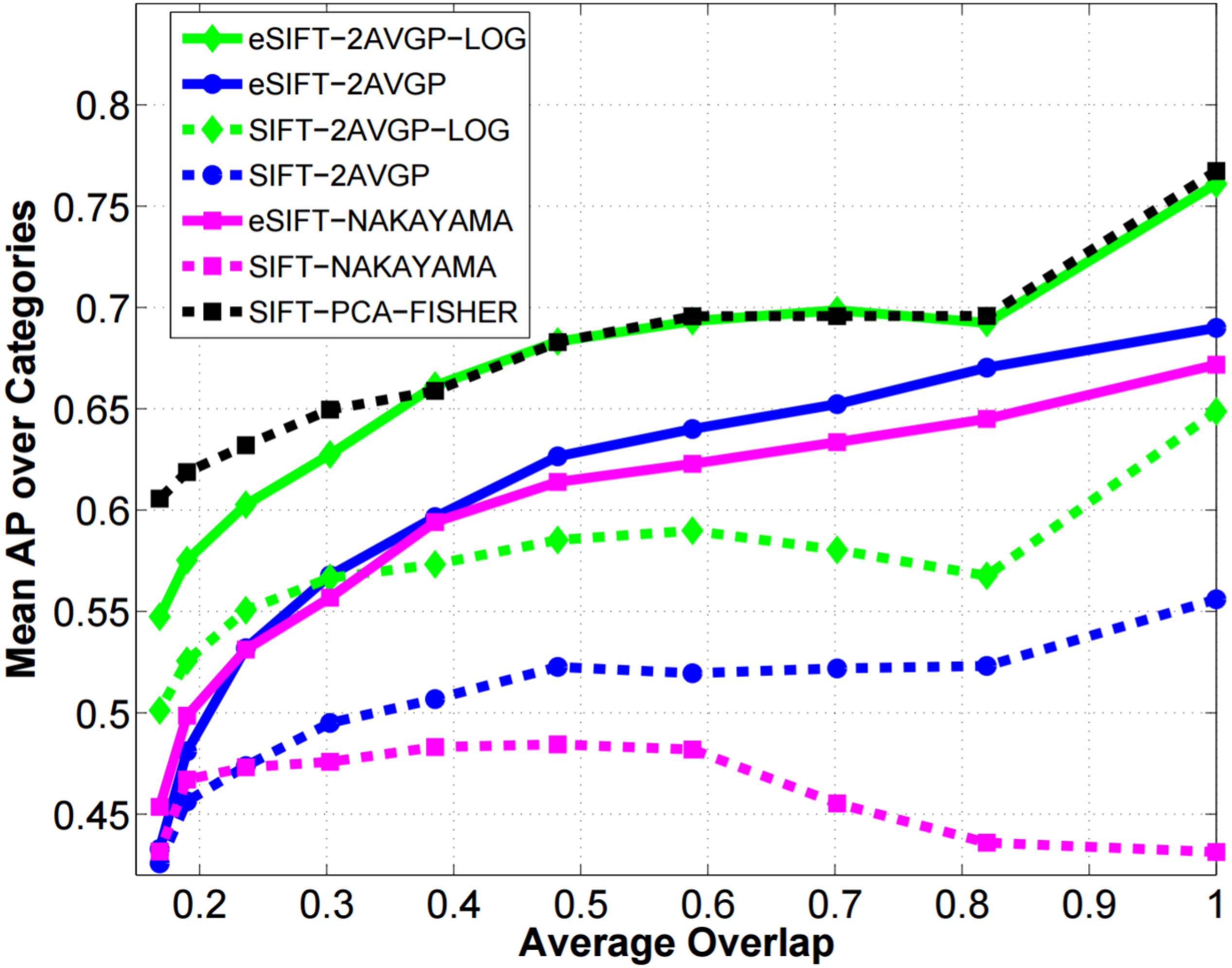
J. Carreira *et al.* Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

Free-form Region Modeling

- SIFT/Enhanced SIFT + $\frac{1}{N}\mathbf{X}\mathbf{X}^T$
- SIFT/Enhanced SIFT + $\log\left(\frac{1}{N}\mathbf{X}\mathbf{X}^T\right)$
- SIFT/Enhanced SIFT + Gaussian-Center Tangent Kernel
- SIFT + Fisher Vector

J. Carreira *et al.* Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

Free-form Region Modeling



Enhanced SIFT + $\log\left(\frac{1}{N}\mathbf{X}\mathbf{X}^T\right)$

Winner of semantic segmentation On Pascal VOC2012

PASCAL2 Pattern Analysis, Statistical Modelling and Computational Learning

J. Carreira *et al.* Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

Free-form Region Modeling




Caltech 101 with Clear Background

SIFT-O2P	eSIFT-O2P	LLC	Fisher Vector
79.2	**80.8**	73.4	77.8

J. Carreira *et al.* Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

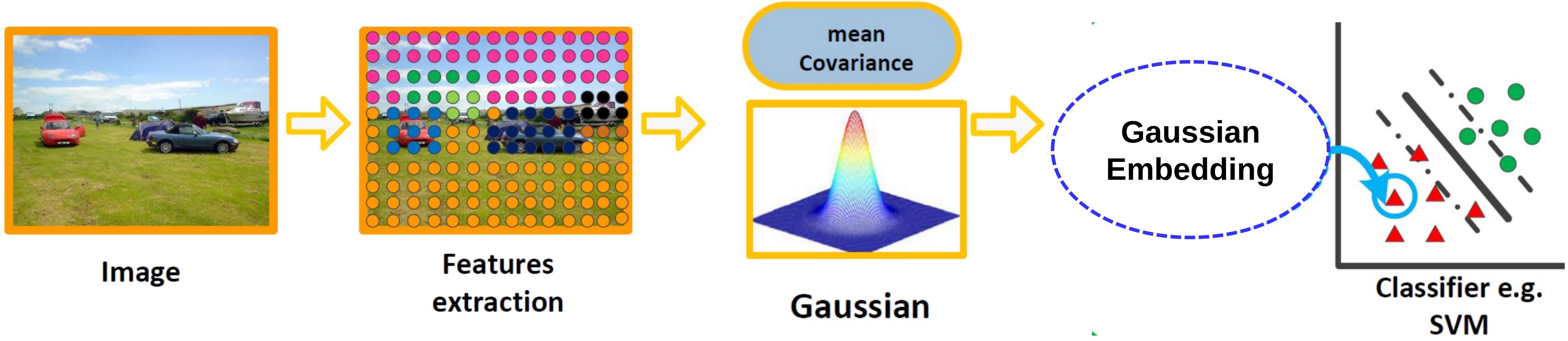
Free-form Region Modeling

1. How about enhanced SIFT + Fisher vector ?

2. Clear Background ?

J. Carreira *et al.* Freeform region description with second-order pooling. *IEEE TPAMI*, 2015.

Whole Image Modeling



- **Enhanced Local (hand-crafted) Features**
- **Modified Gaussian Embedding**

Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Whole Image Modeling

■ Enhanced Local (hand-crafted) features

- SIFT [IJCV 03]
- Enhanced SIFT [ECCV 12] (Color + Location + Filters ……)
- L^2EMG [TPAMI 17]
- Enhanced L^2EMG

Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Whole Image Modeling

Modified Gaussian Embedding

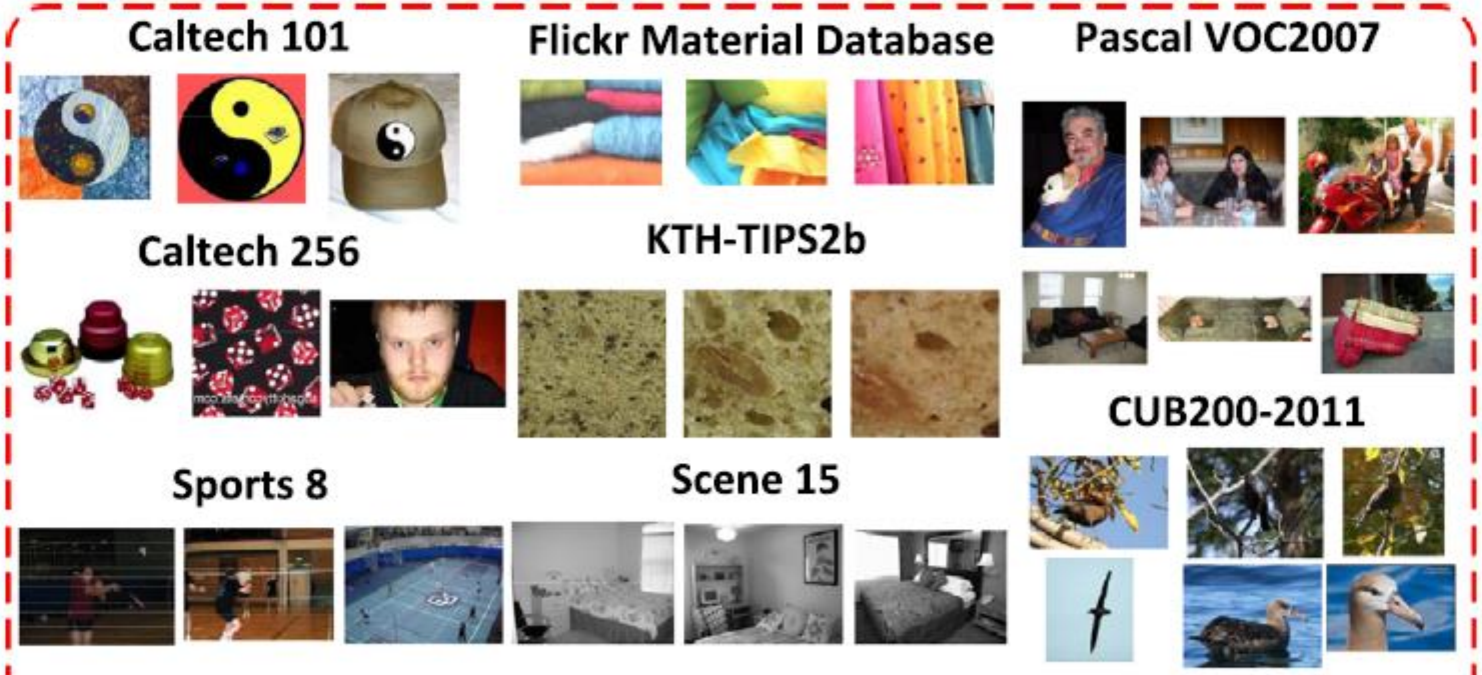
$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) \overset{\pi}{\mapsto} \mathbf{A} = \begin{bmatrix} \mathbf{P} & \boldsymbol{\mu} \\ \mathbf{0}^T & 1 \end{bmatrix} \overset{\gamma}{\mapsto} \mathbf{S} = \begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T & \boldsymbol{\mu} \\ \boldsymbol{\mu}^T & 1 \end{bmatrix} \overset{\log}{\mapsto} \log\left(\begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T & \boldsymbol{\mu} \\ \boldsymbol{\mu}^T & 1 \end{bmatrix}\right)$$

$$\Downarrow$$

$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) \overset{\pi(\beta)}{\mapsto} \mathbf{A}(\beta) = \begin{bmatrix} \mathbf{P} & \beta\boldsymbol{\mu} \\ \mathbf{0}^T & 1 \end{bmatrix} \overset{\gamma(\rho)}{\mapsto} \mathbf{S}(\beta,\rho) = \begin{bmatrix} \boldsymbol{\Sigma}^\rho + \beta^2\boldsymbol{\mu}\boldsymbol{\mu}^T & \beta\boldsymbol{\mu} \\ \beta\boldsymbol{\mu}^T & 1 \end{bmatrix} \overset{\log}{\mapsto} \log\left(\begin{bmatrix} \boldsymbol{\Sigma}^\rho + \beta^2\boldsymbol{\mu}\boldsymbol{\mu}^T & \beta\boldsymbol{\mu} \\ \beta\boldsymbol{\mu}^T & 1 \end{bmatrix}\right)$$

Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Whole Image Modeling



	Caltech 101	Caltech 256	KTH-TIPS2b	FMD	VOC2007	CUB200-2011	Scene 15	Sports 8
FV	83.7	50.1	71.3	58.9	61.8	27.3	89.4	91.3
Our CLM	86.3	53.6	75.2	64.2	61.7	36.0	89.4	91.5

Fig. 1. Some example images and accuracy comparison (in %) between Fisher vector (FV) and our codebookless model (CLM) on various image databases.

CLM $\geq$ Fisher Vector

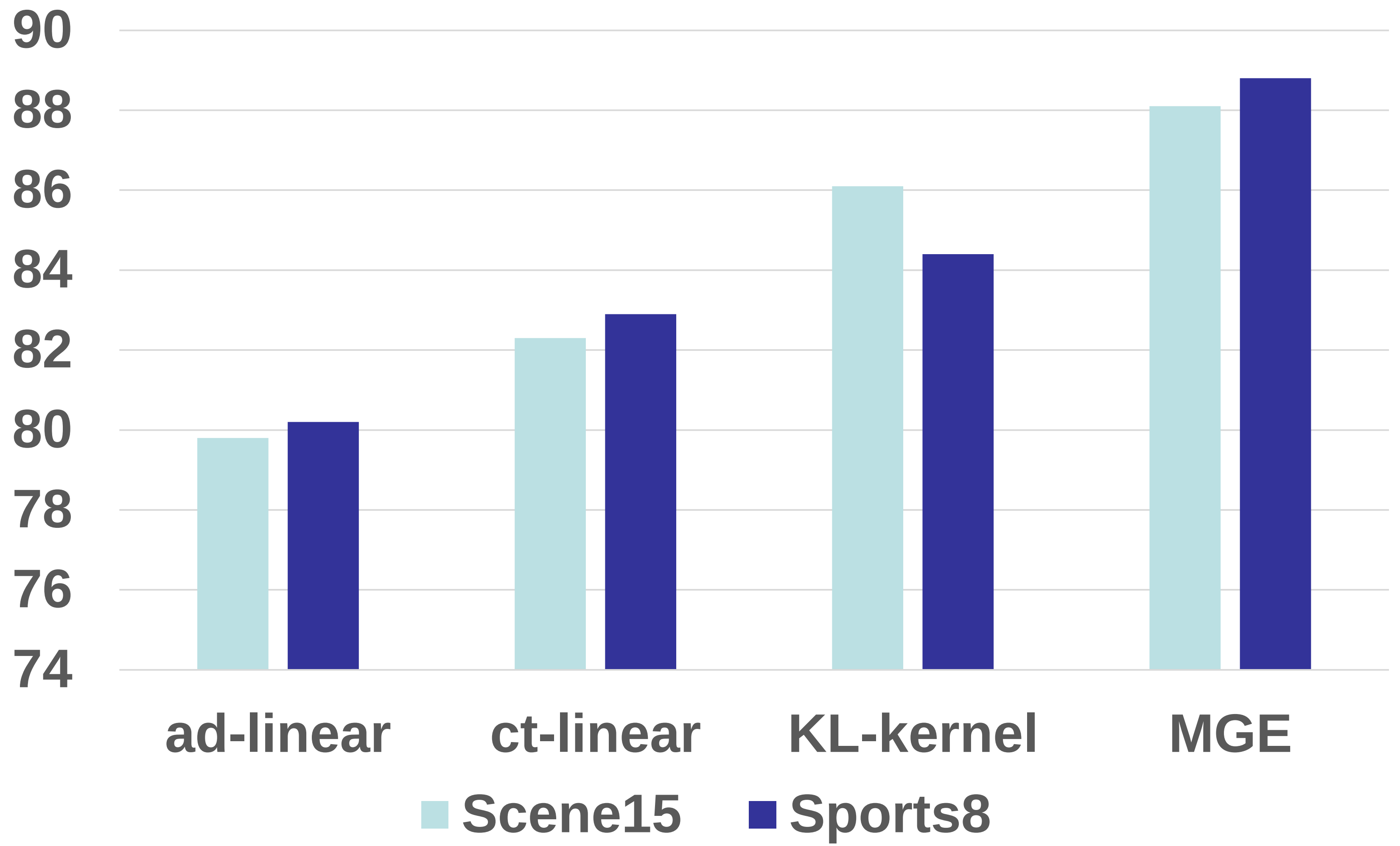
Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Whole Image Modeling

	Caltech 101	Caltech 256	VOC2007	CUB200-2011	FMD	KTH-TIPS-2b	Scene15	Sports8
FV+SIFT	80.87	47.47	**61.8**	25.8	58.37	69.37	88.17	**91.37**
FV+eSIFT	83.77	50.17	**60.8**	27.3	58.9	71.37	89.47	**90.47**
CLM+SIFT	84.97	48.97	55.8	18.6	51.67	71.87	88.17	88.87
CLM+eSIFT	**86.37**	**53.67**	60.4	28.1	57.77	**75.27**	**89.47**	**91.57**
CLM+L^{2}EMG	82.57	48.67	56.6	19.1	62.47	72.27	88.37	88.37
CLM+eL^{2}EMG	84.77	53.27	**61.7**	**28.6**	**64.27**	73.67	89.27	90.77

Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

Whole Image Modeling



$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \begin{bmatrix} \boldsymbol{\Sigma}^{\rho} + \beta^2 \boldsymbol{\mu}\boldsymbol{\mu}^T & \beta\boldsymbol{\mu} \\ \beta\boldsymbol{\mu}^T & 1 \end{bmatrix}$$

Wang *et al.* Towards Effective Codebookless Model for Image Classification. Pattern Recognition, 2016

BoVW *VS.* CLM – Summary

- Higher-order CLM (e.g., single Gaussian) is a very competitive alternative to BoVW model

- Efficient and effective usage of geometry of higher-order CLM is a key issue

- Higher-order CLM is more **sensitive** to **local descriptors** than BoVW model

Context

1
- Higher-order Statistics in Bag-of-Visual-Words (BoVW)

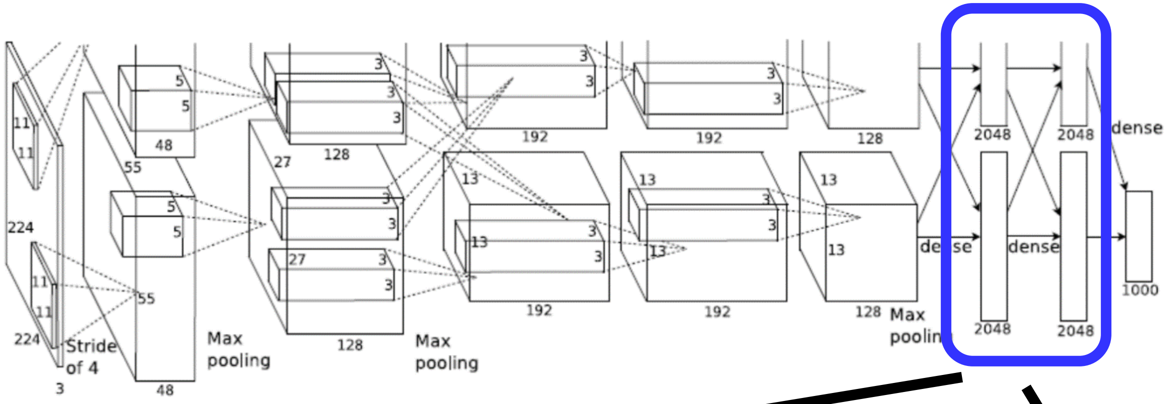
2
- Higher-order Statistics in Codebookless Model (CLM)

3
- Bag-of-Visual-Words *vs.* Codebookless Model

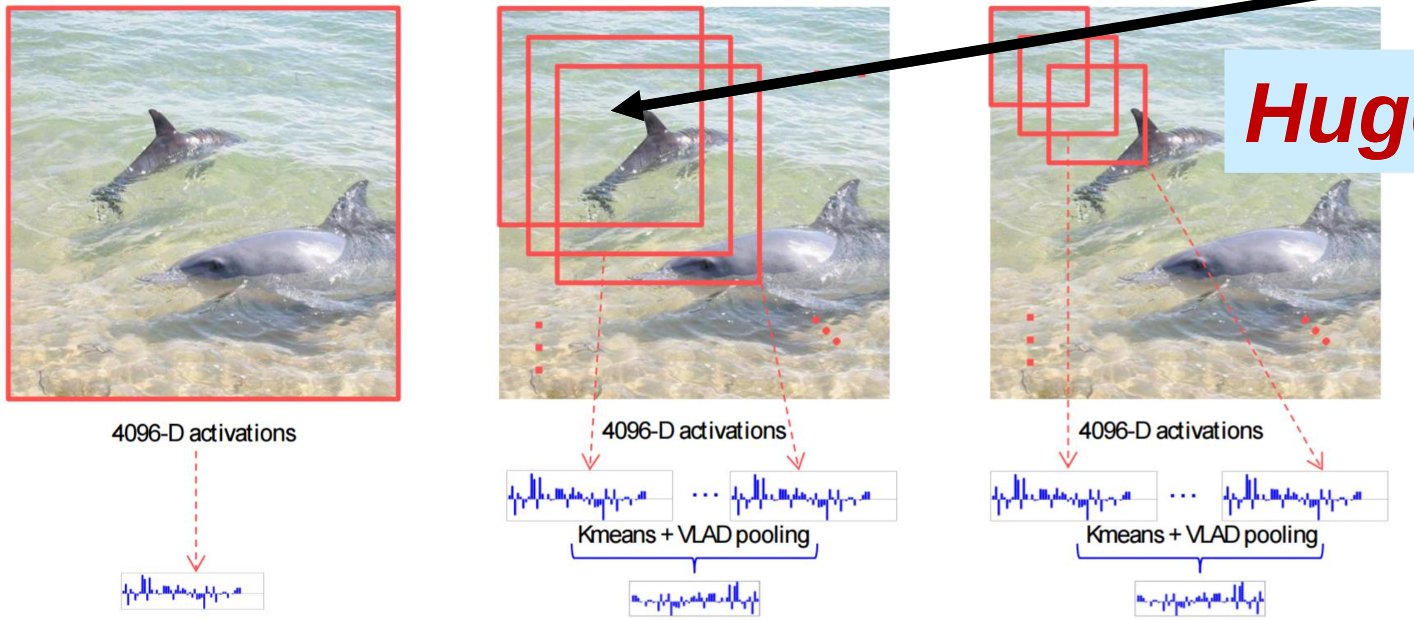
4
- **Higher-order Statistical Models Meet Deep Features**

Coding for Deep Features

AlexNet
[NIPS 2012]



MOP-CNN
[ECCV 2014]



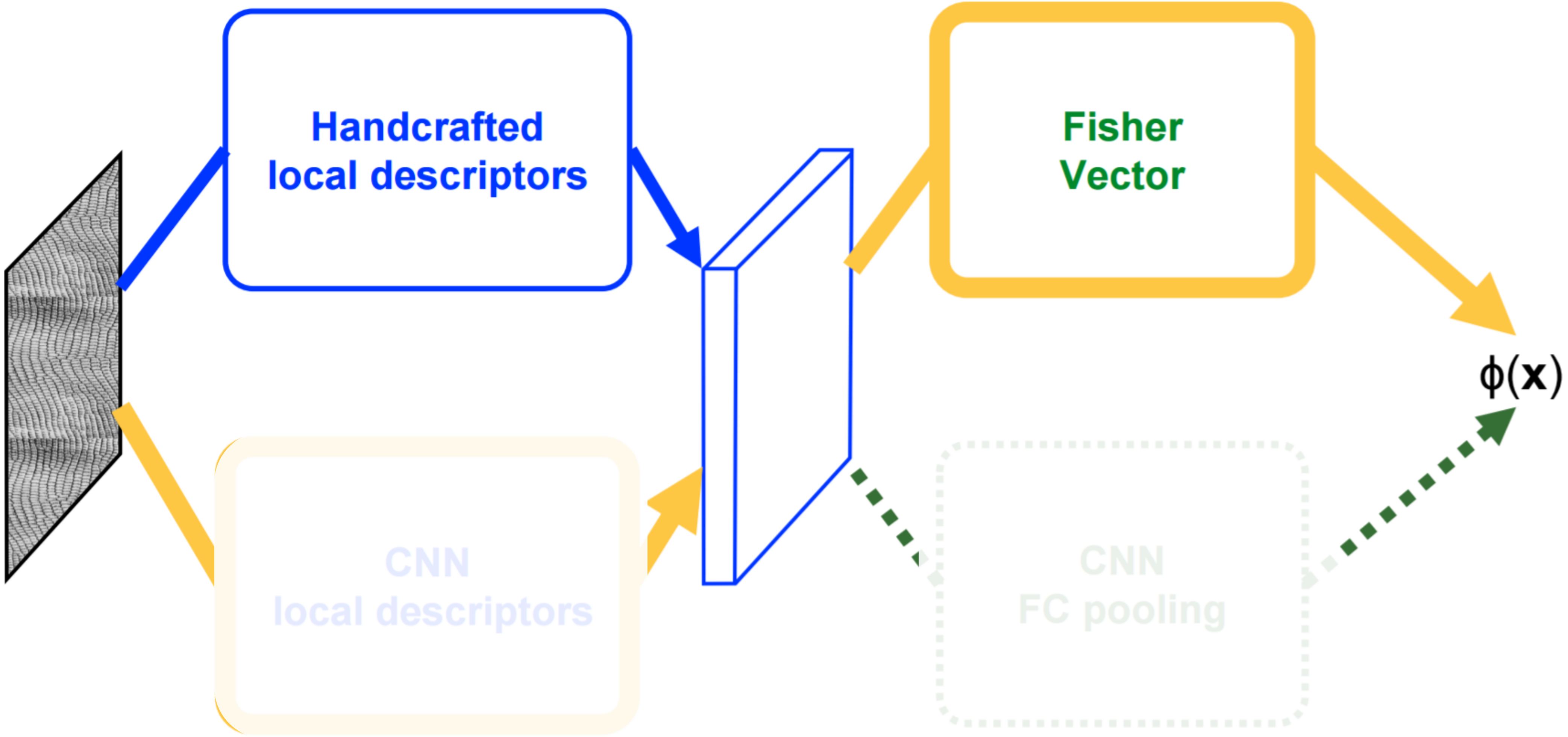
(a) level1: global activation (b) level2: pooled features (c) level3: pooled features

Huge Computational burden!

$$\sum_{\mathbf{x}_i \in \mathcal{I}} (\mathbf{x}_i - \mathbf{B}\mathbf{u}_i^*)\mathbf{u}_i^{*\top}$$

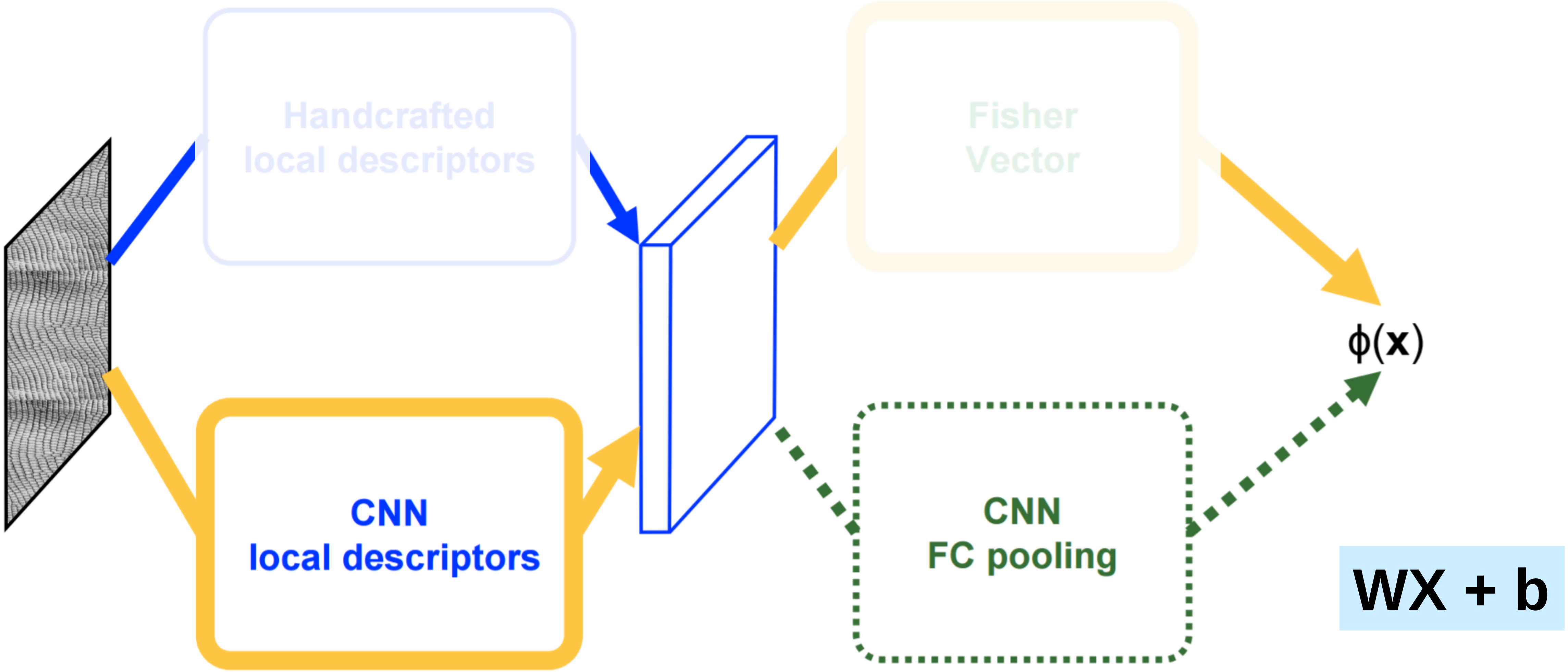
SCFVC [NIPS2014]

FV-CNN



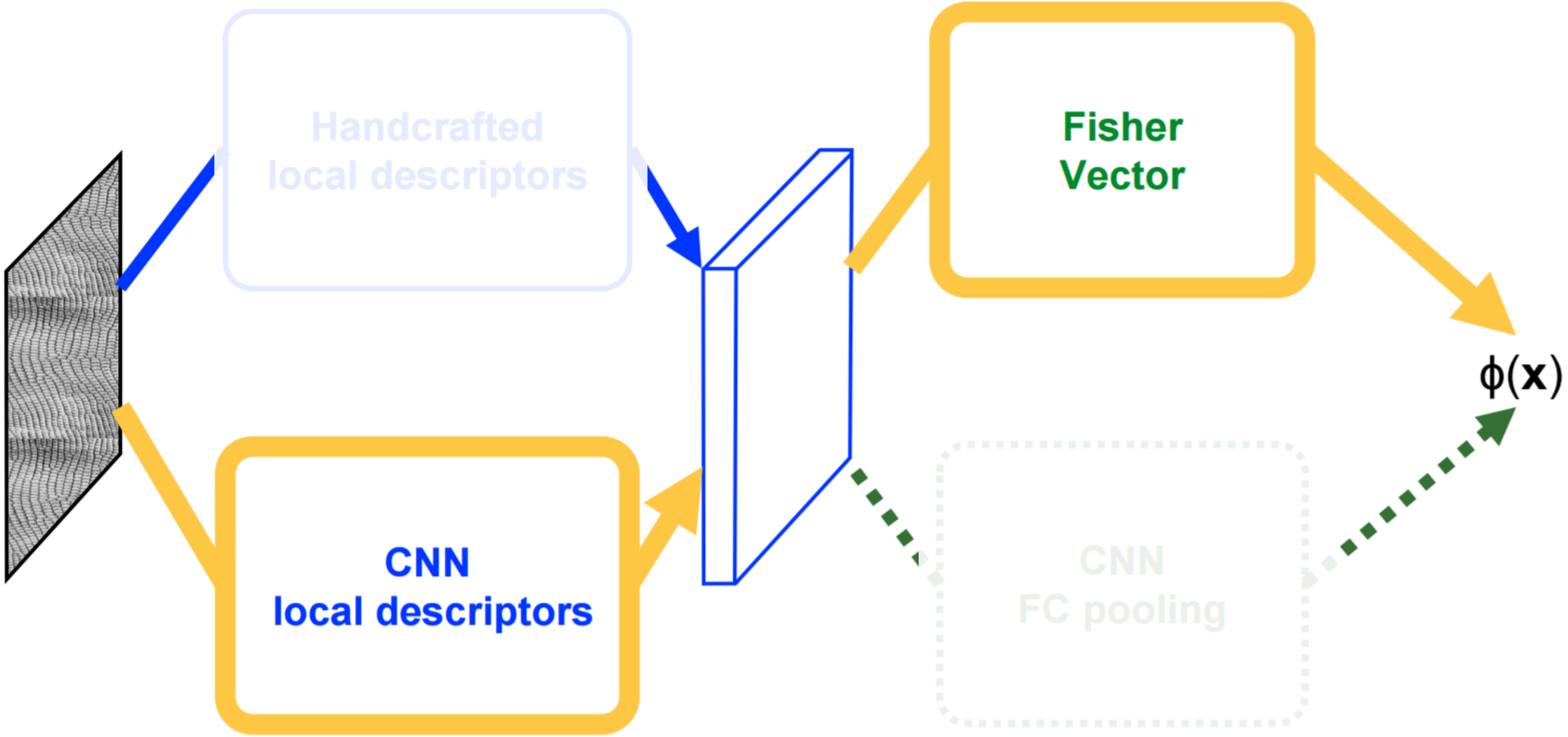
M. Cimpoi et al. Deep filter banks for texture recognition and segmentation. In *CVPR*, 2015.

FV-CNN



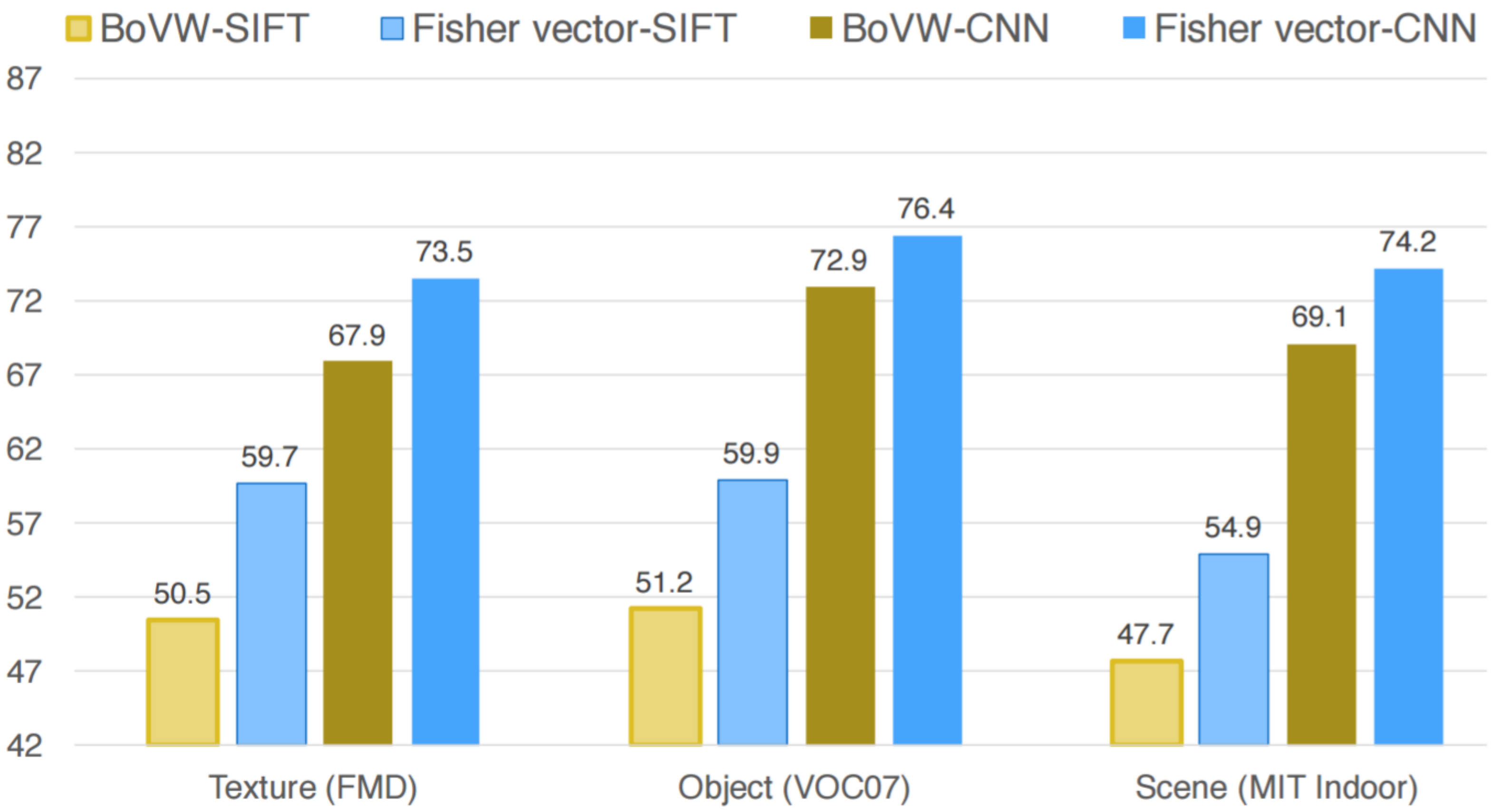
M. Cimpoi et al. Deep filter banks for texture recognition and segmentation. In *CVPR*, 2015.

FV-CNN



M. Cimpoi et al. Deep filter banks for texture recognition and segmentation. In *CVPR*, 2015.

FV-CNN

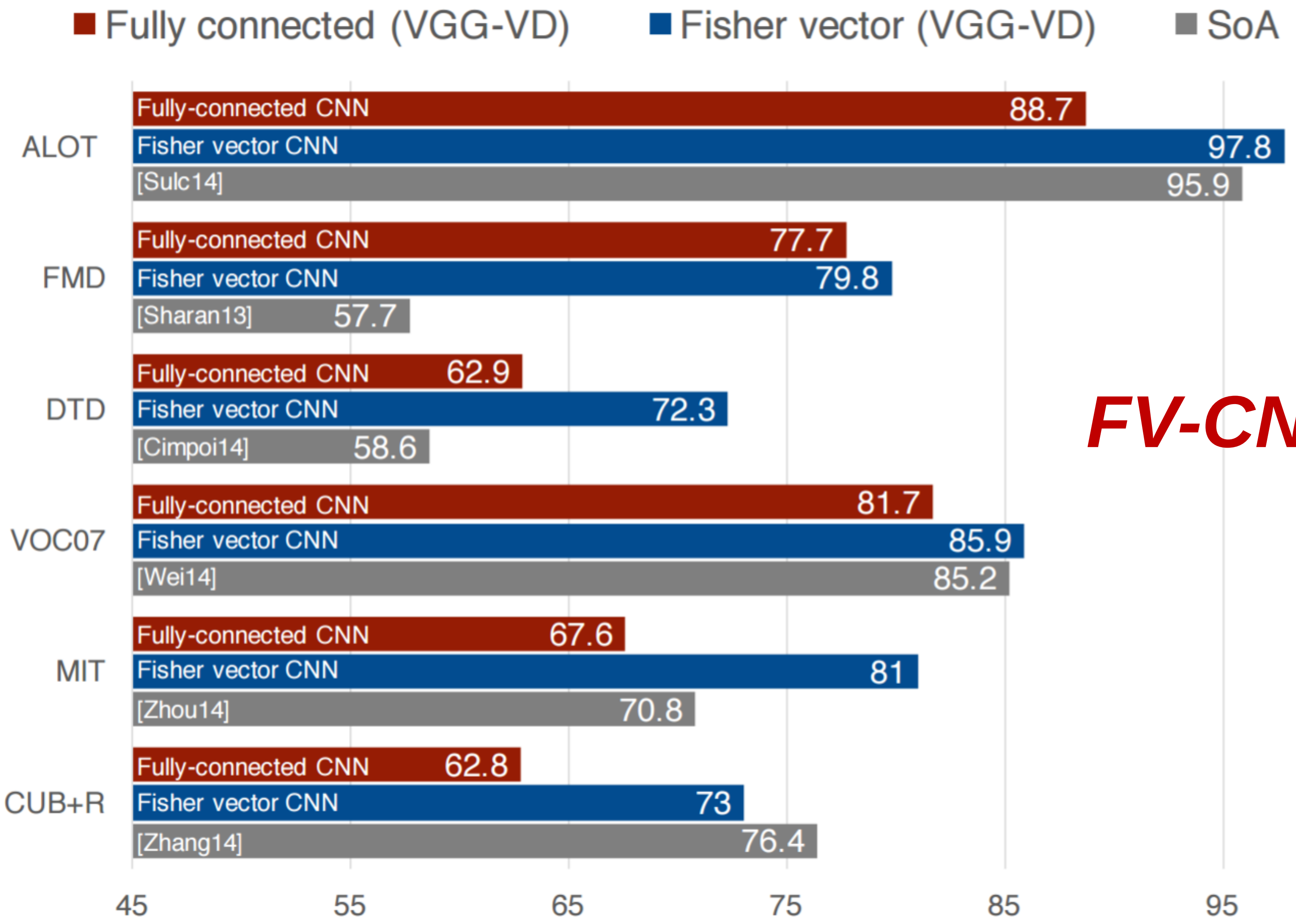


Finding 1) BoVW < FV

Finding 2) SIFT < CNN

M. Cimpoi et al. Deep filter banks for texture recognition and segmentation. In *CVPR*, 2015.

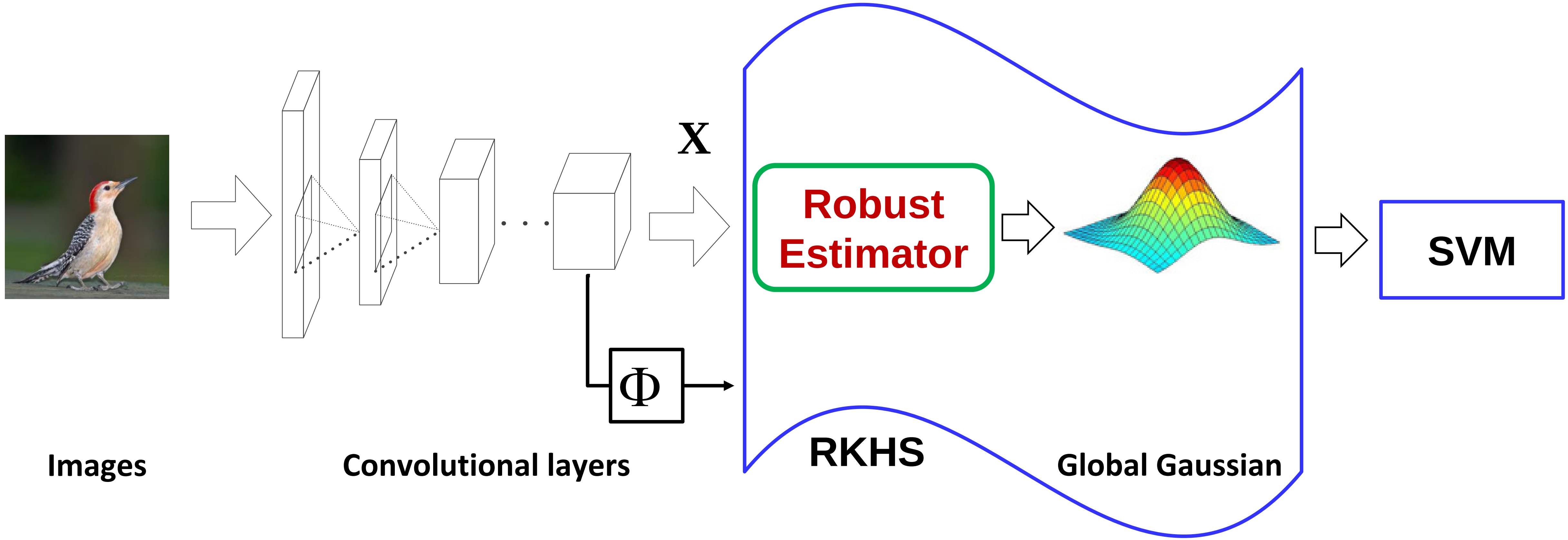
FV-CNN



FV-CNN >> FC Pooling !

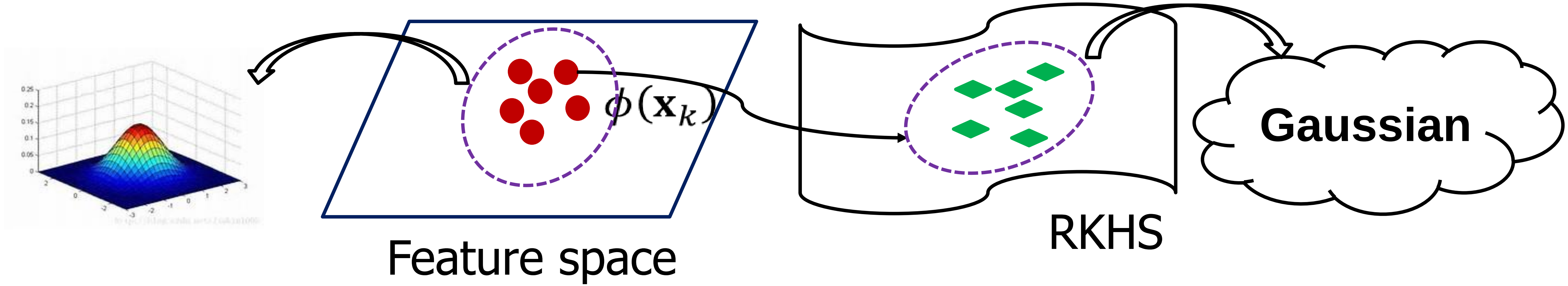
M. Cimpoi … **n *CVPR*, 2015.**

RIAD-G



Wang et al. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

RIAD-G



$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{k=1}^{N}\boxed{\phi(\mathbf{x}_k)}, \quad \hat{\mathbf{S}} = \frac{1}{N-1}\Phi(\mathbf{X})\mathbf{J}\Phi(\mathbf{X})^T.$$

Hellinger's and $\mathcal{X}^2$ Kernel [TPAMI 11]

Wang et al. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

RIAD-G

$$p(\mathbf{x}) = |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$\min_{\boldsymbol{\Sigma}} \frac{N}{2}\log|\boldsymbol{\Sigma}| + \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})$$

$$\Downarrow$$

Not Robust!

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^T$$

Classical MLE

$$\min_{\widehat{\boldsymbol{\Sigma}}} \log|\widehat{\boldsymbol{\Sigma}}| + \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{S}}) + \boxed{\alpha D_{\mathrm{vN}}(\mathbf{I}, \widehat{\boldsymbol{\Sigma}})}$$

$$\Downarrow$$

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{U}}\mathrm{diag}(\lambda_k)\widehat{\mathbf{U}}^T,$$

$$\lambda_k = \sqrt{\left(\frac{1-\alpha}{2\alpha}\right)^2 + \frac{\delta_k}{\alpha}} - \frac{1-\alpha}{2\alpha}$$

vN-MLE

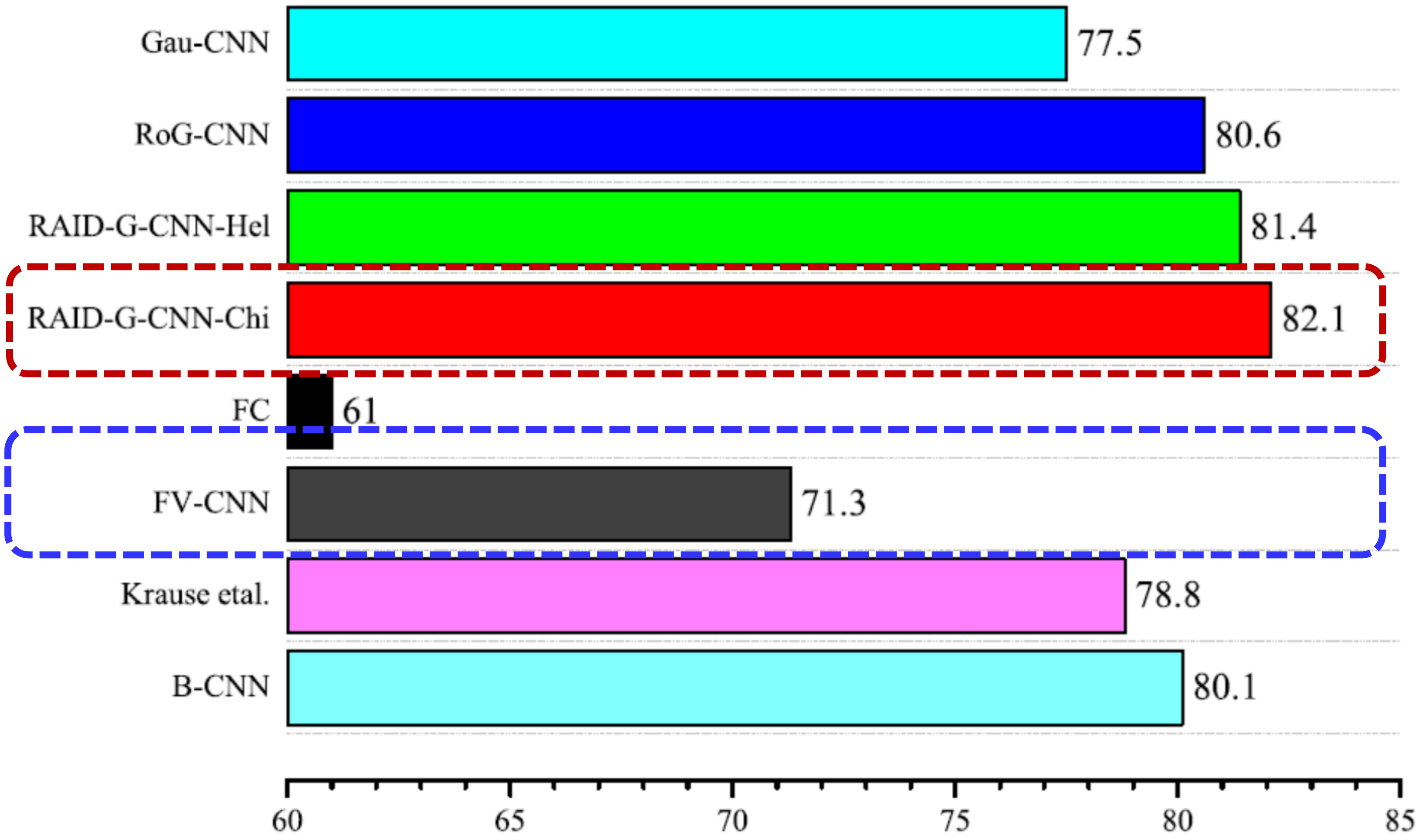
Wang et al. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Comparison

Methods	FMD	UIUC Material	KTH-TIPS 2b	DTD	Open Surfaces
COV-CNN	80.2 ± 1.1	80.5 ± 3.6	76.7 ± 2.8	70.1 ± 1.2	55.0
Gau-CNN	81.3 ± 1.4	81.7 ± 2.9	77.5 ± 2.4	70.5 ± 1.5	55.7
RoG-CNN	83.6 ± 1.6	84.5 ± 1.8	79.5 ± 1.5	73.9 ± 1.1	58.9
RAID-G-CNN-Hel	84.4 ± 1.3	85.7 ± 2.1	80.4 ± 1.2	75.8 ± 1.4	60.3
RAID-G-CNN-Chi	$\mathbf{84.9 \pm 1.4}$	$\mathbf{86.3 \pm 2.9}$	81.3 ± 1.6	$\mathbf{76.4 \pm 1.1}$	**61.1**
FC [12]	77.4 ± 1.8	75.9 ± 2.3	75.4 ± 1.5	62.9 ± 0.8	43.4
FV-CNN [12]	79.8 ± 1.8	80.5 ± 2.7	$\mathbf{81.8 \pm 2.5}$	72.3 ± 1.0	59.5
FC + FV-CNN* [12]	82.4 ± 1.5	82.6 ± 2.1	81.1 ± 2.4	74.7 ± 1.0	60.9
State-of-the-art I	60.6 [42]	60.1 [18]	70.7 ± 1.6 [16]	61.2 ± 1.0 [40]	39.8 [40]
State-of-the-art II	66.5 ± 1.5 [4]	66.6 ± 3.1 [22]	77.3 ± 2.3 [11]	66.7 ± 0.9 [11]	-

Wang et al. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Comparison



Birds CUB-200-2011

Wang et al. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Summary

- Deep CNN features significantly improve higher-order models
- Higher-order models can significantly improve FC pooling
- Higher-order CLM outperforms Higher-order BoVW using deep features
- Robust estimation is important for higher-order CLM under deep CNNs

Take home message

- Higher-order statistics plays a key role in classical modeling methods: BoVW and CLM

- Comparison with higher-order CLM and higher-order BoVW model using both hand-crafted features and deep features

- It is useful to combine higher-order statistics modeling with pre-trained deep CNNs in a separated manner

Question ?

Can we integrate higher-order CLM into deep CNN architectures in an end-to-end learning manner for further improvement?

Our Related Publications

1. Peihua Li, Qilong Wang, Hui Zeng and Lei Zhang. Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification. **IEEE TPAMI** 39(4): 803-817, **2017**.
2. Peihua Li, Hui Zeng, Qilong Wang, Simon C. K. Shiu, Lei Zhang. High-order Local Pooling and Encoding Gaussians over A Dictionary of Gaussians. **IEEE TIP**, **2017**
3. Qilong Wang, Peihua Li, Wangmeng Zuo, Lei Zhang. Towards Effective Codebookless Model for Image Classification. **Pattern Recognition** 59: 63-71, **2016**.
4. Qilong Wang, Peihua Li, Wangmeng Zuo, Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition. 29th IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), **2016**.
5. Peihua Li, Xiaoxiao Lu, Qilong Wang. From Dictionary of Visual Words to Subspaces: Locality-constrained Affine Subspace Coding. 28th IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), **2015**.
6. Peihua Li, Qilong Wang, Lei Zhang. A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models. 14th IEEE International Conference on Computer Vision (**ICCV**), **2013**.

Thank you!