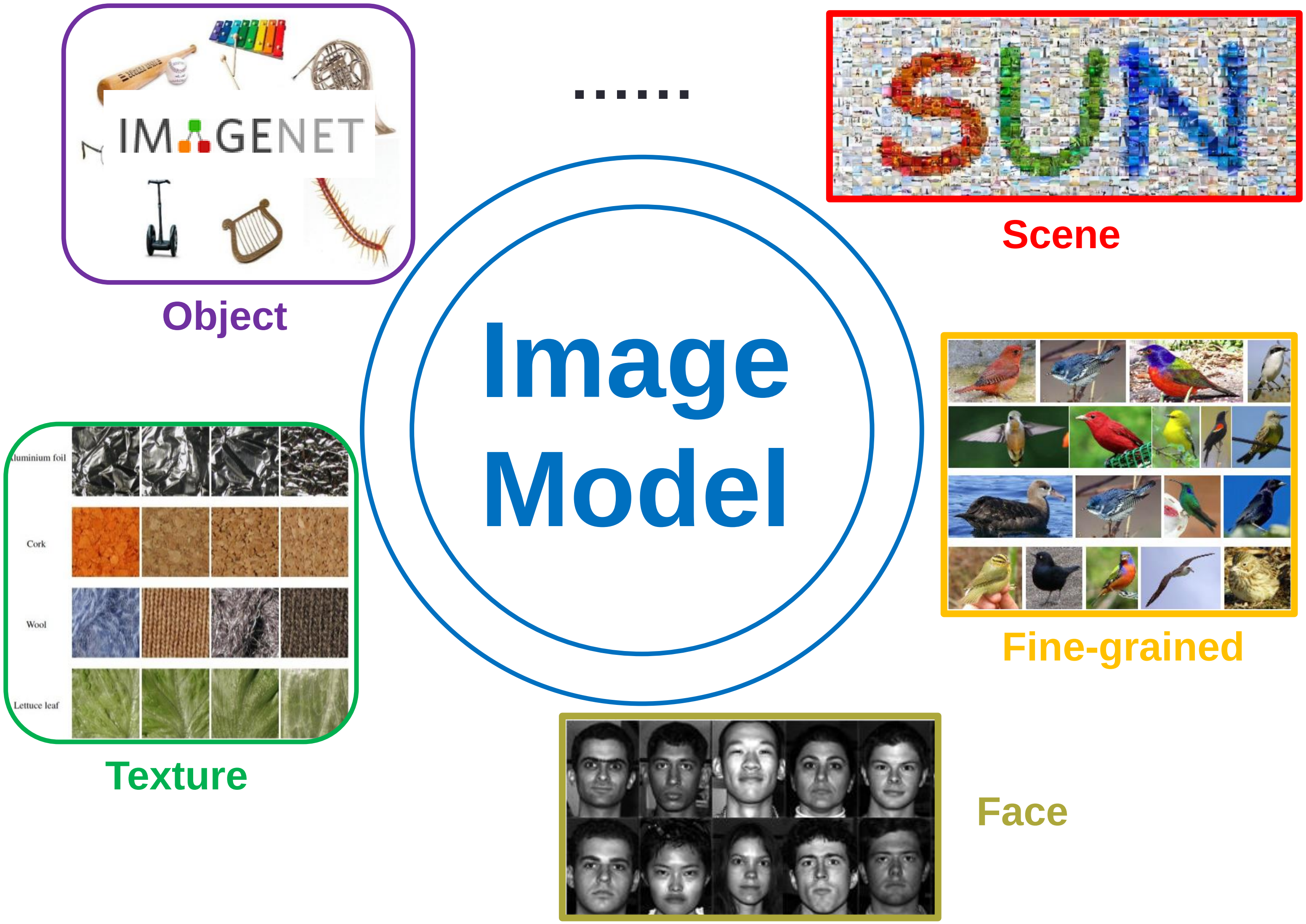
Vision by Learning Group at DUT

Codebook-free Single Gaussian for Image Classification

Qilong Wang (王旗龙)
Dalian University of Technology

http://ice.dlut.edu.cn/PeihuaLi/

Image Model for Classification



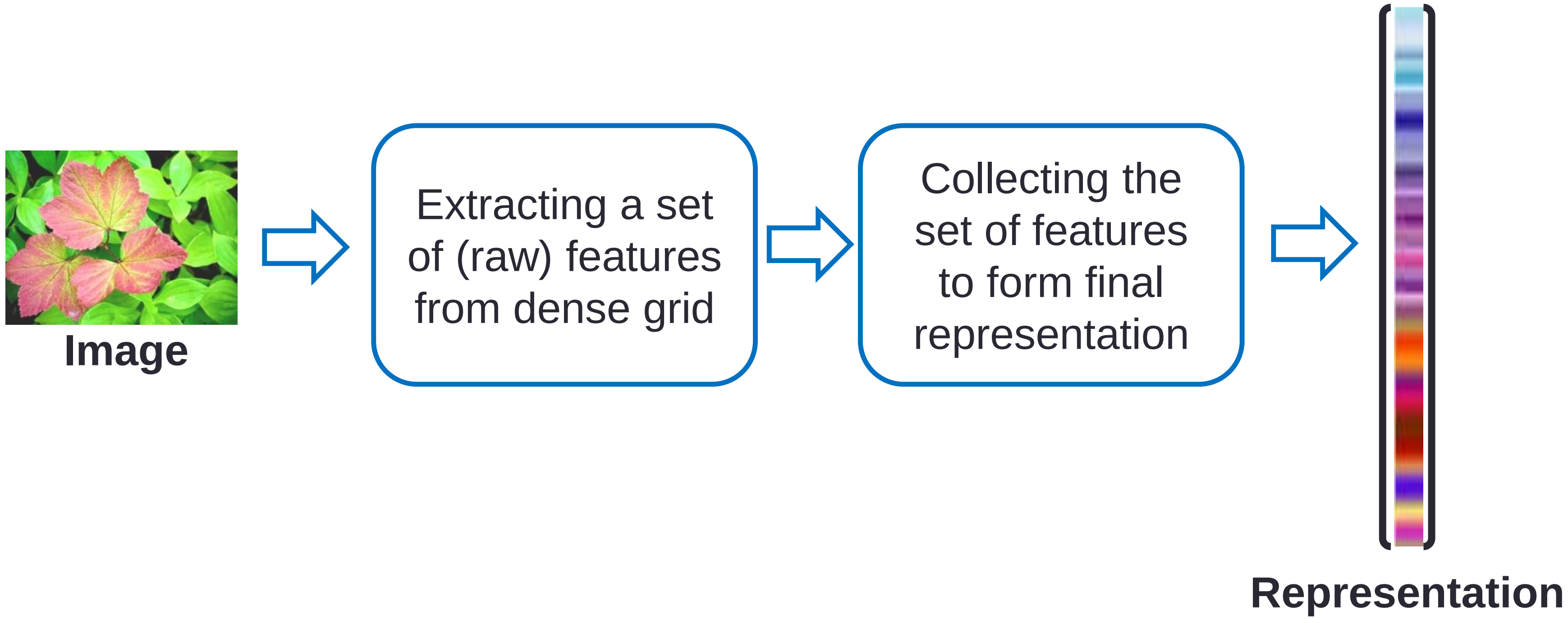
Outline

- ***Modeling Methods in Image Classification***
- ***Towards Effective Codebook-free Model***
- ***Robust Approximate Infinite Dimensional Gaussian***
- ***Future Work and Conclusion***

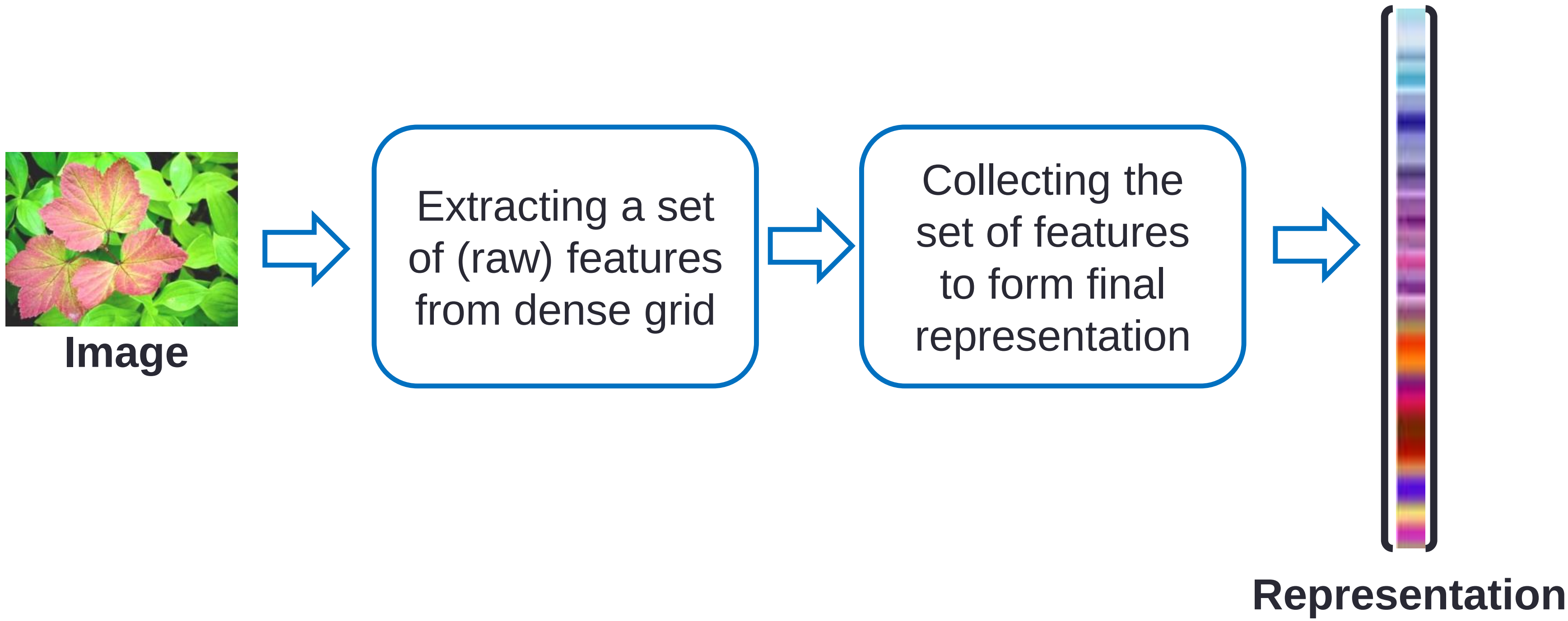
Outline

- ***Modeling Methods in Image Classification***
- ***Towards Effective Codebook-free Model***
- ***Robust Approximate Infinite Dimensional Gaussian***
- ***Future Work and Conclusion***

Modeling Methods in Image Classification

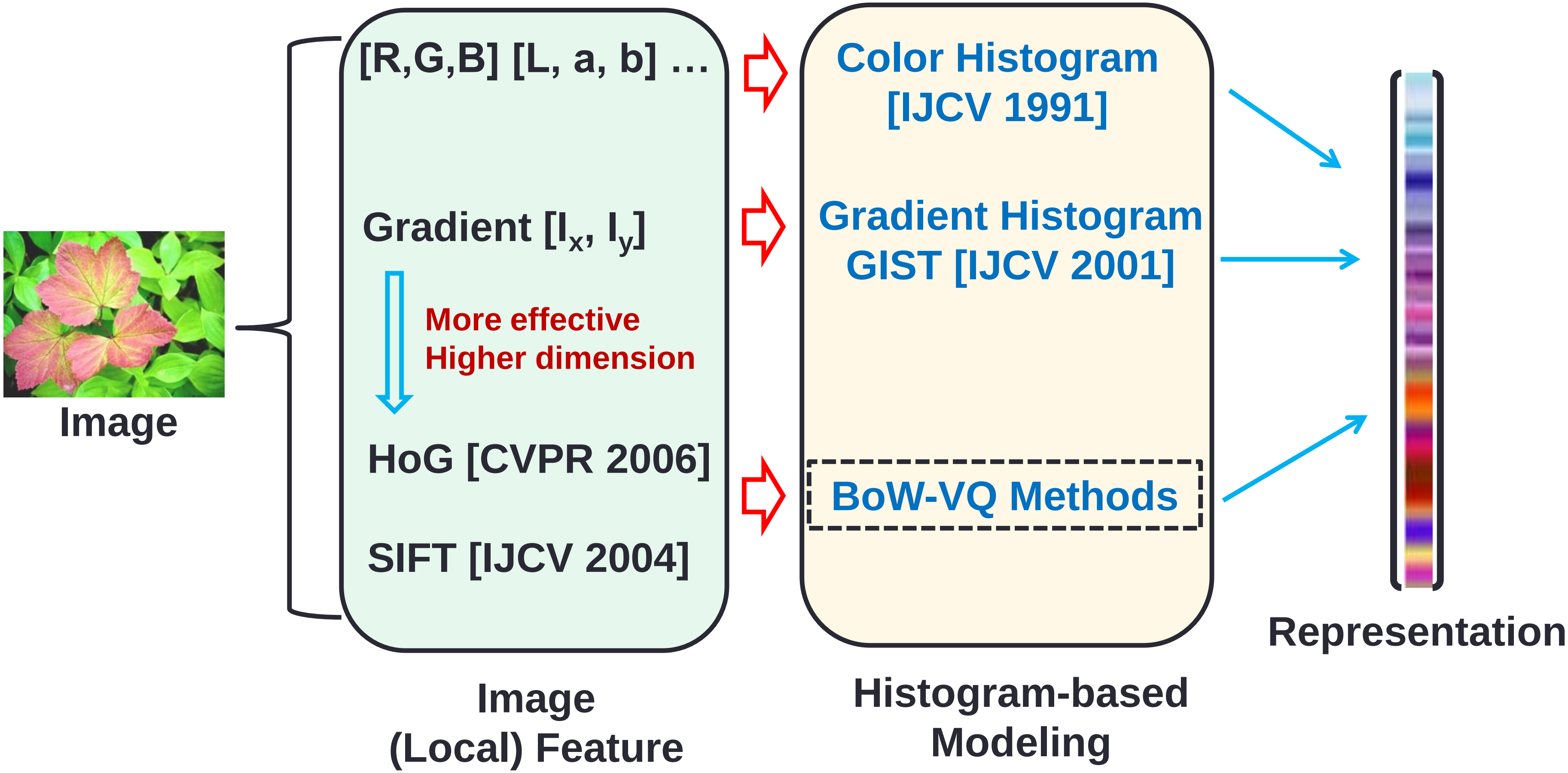


Modeling Methods in Image Classification

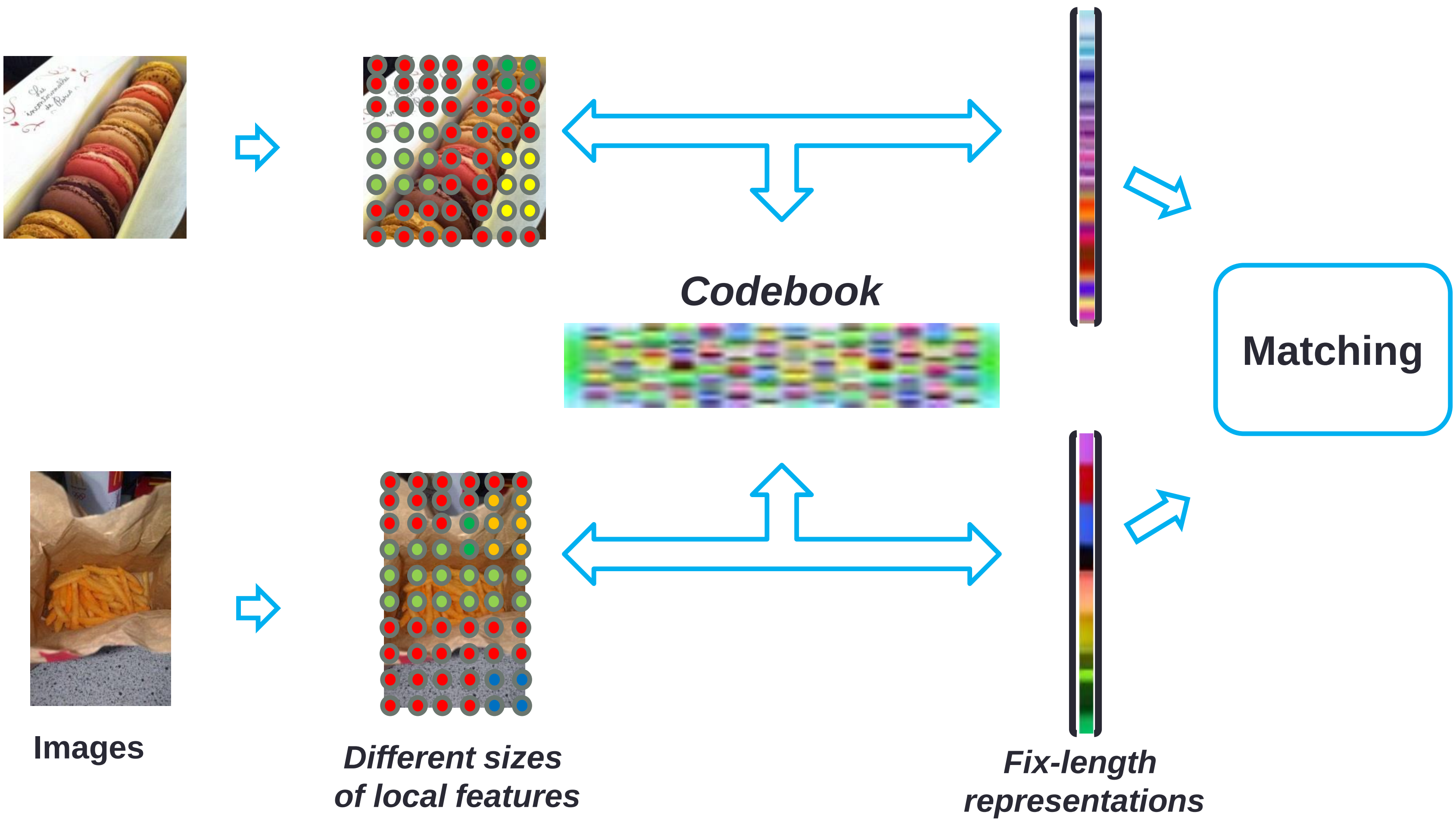


- ☐ **Histogram(Codebook)-based Modeling Methods**
- ☐ **Codebook-free Modeling Methods**

Histogram-based Modeling Methods



Histogram of HD Local Feature – BoW

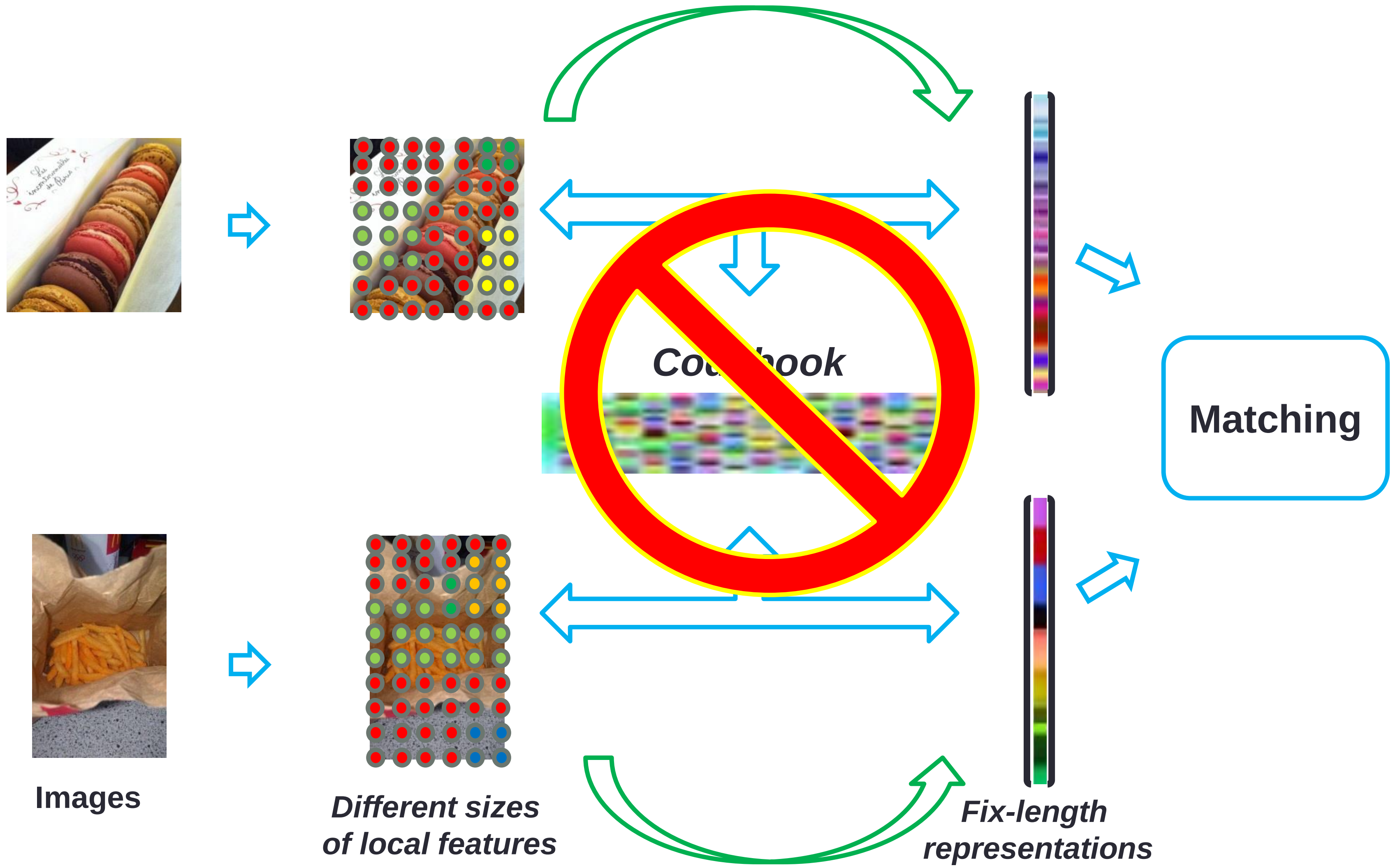


Limitations of BoW

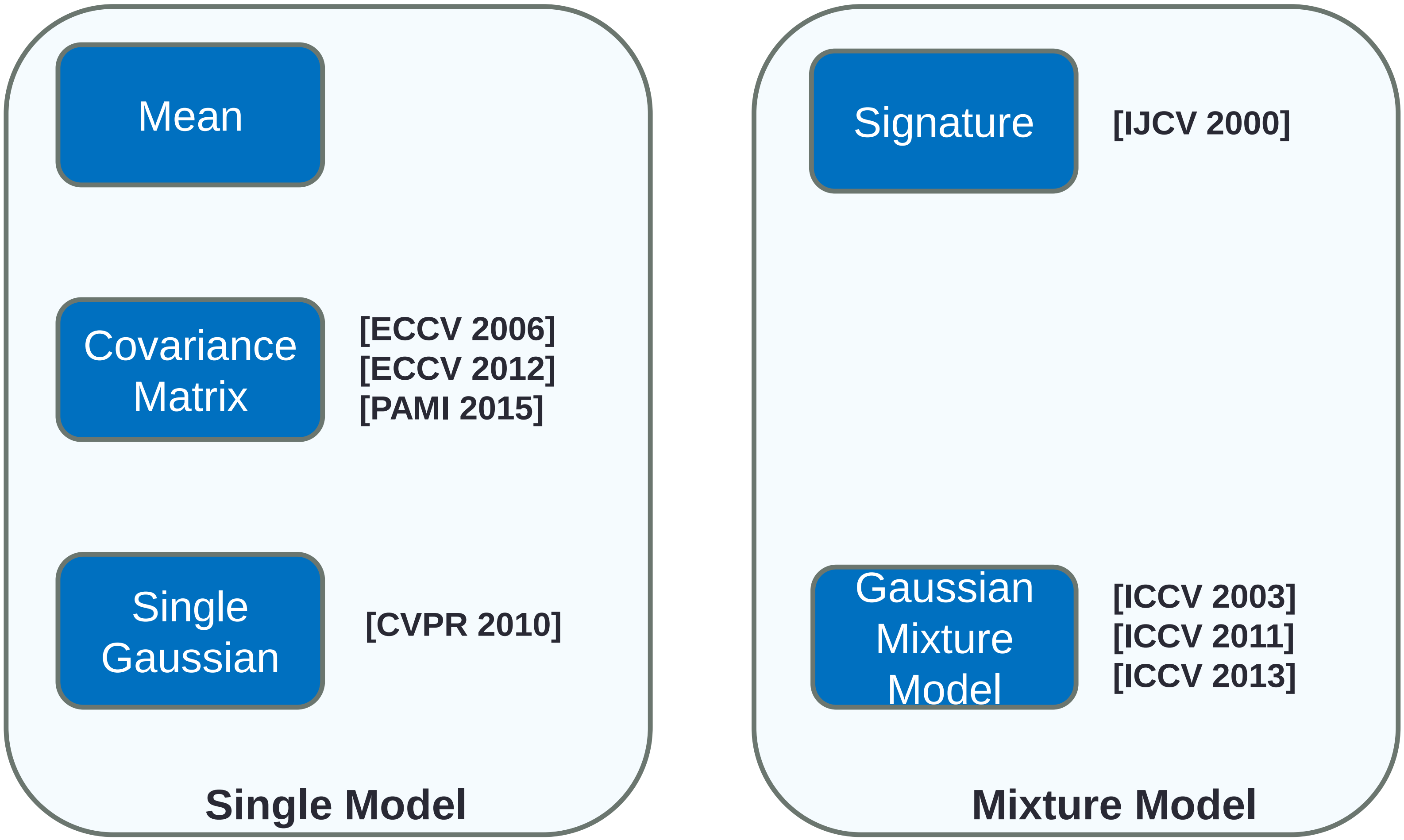
- The codebook brings quantization error. [Boiman et al. CVPR08]
  - Soft-assignment coding methods
    - Visual Word Ambiguity [PAMI10], SC [CVPR 09], LLC[CVPR10],LSAC [ICCV 11]
  - Dictionary enhancement
    - Huge size of dictionary [PAMI15], GMM [IJCV13], Affine subspace [CVPR15] and DL.
  - Usage of first order and second order information
    - VLAD[CVPR10], SV[ECCV10], FV[IJCV13], E-VLAD[ECCV14], LASC[CVPR15].

- An all-purpose codebook is unavailable.
  - It is difficult to handle online problem, e.g., increasing number of classes.

Usage of Codebook-free Model

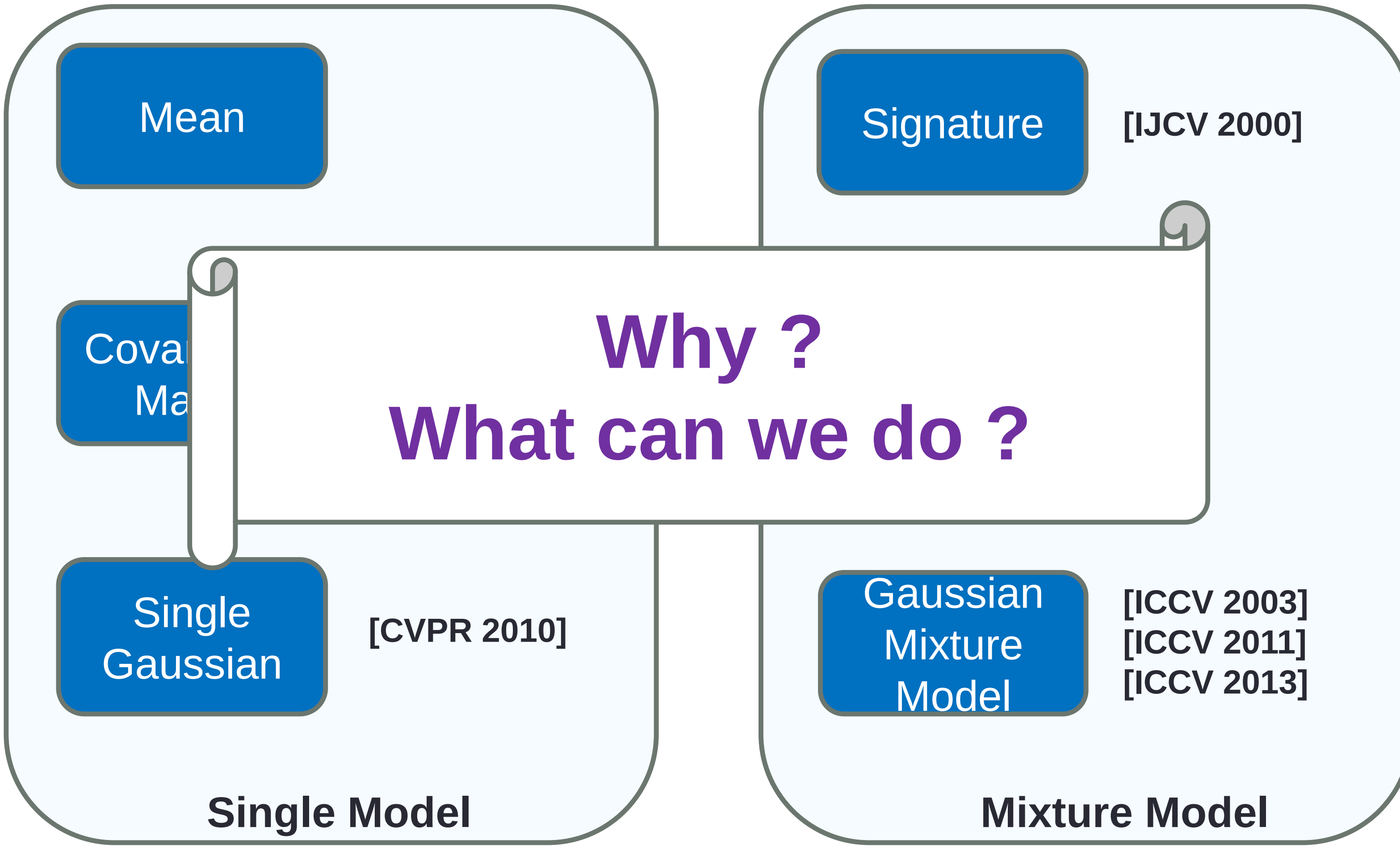


Codebook-free Models



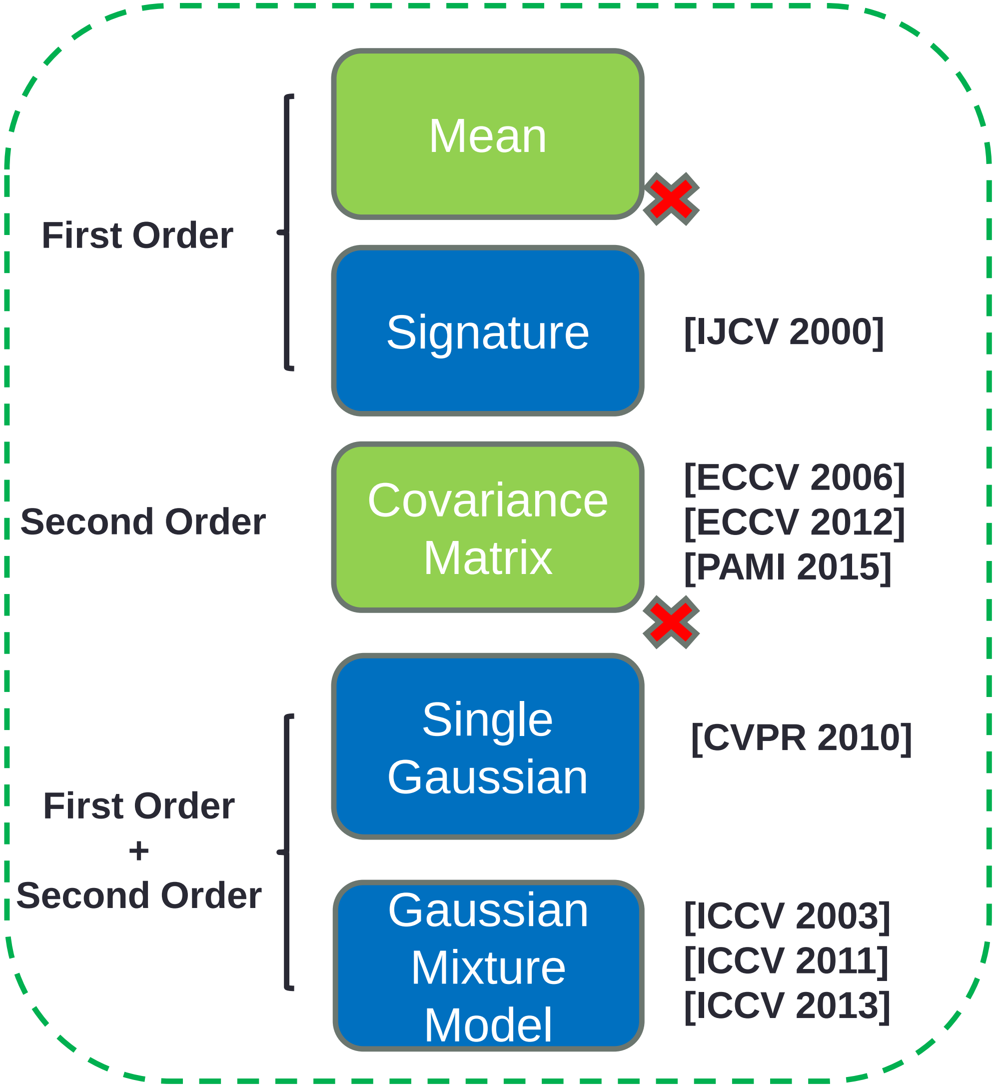
Above models showed underperformances than BoW model for image classification.

Codebook-free Models



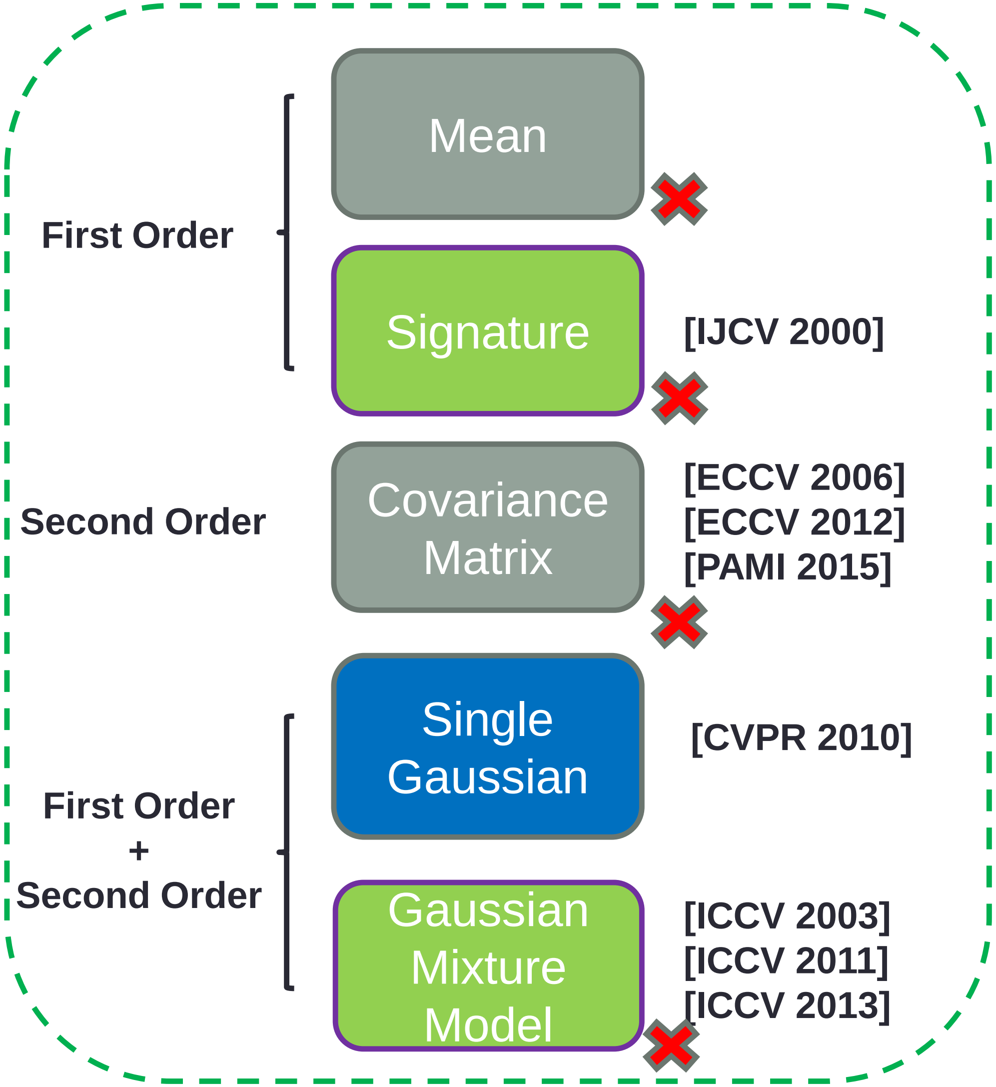
Above models showed underperformances than BoW model for image classification.

Selection of Codebook-free Model



Combination of first and second order brings better performances.

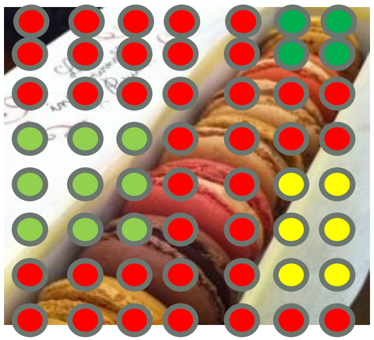
Selection of Codebook-free Model



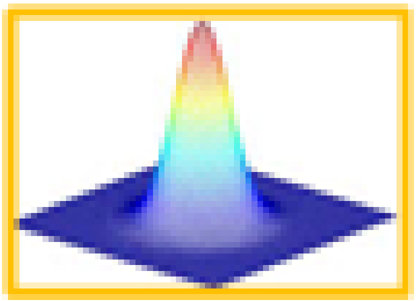
1. Cross-Bin metric is needed.
2. They are difficult to model high dimensional features.

Codebook-free Single Gaussian for Image Modelling

Image



Features

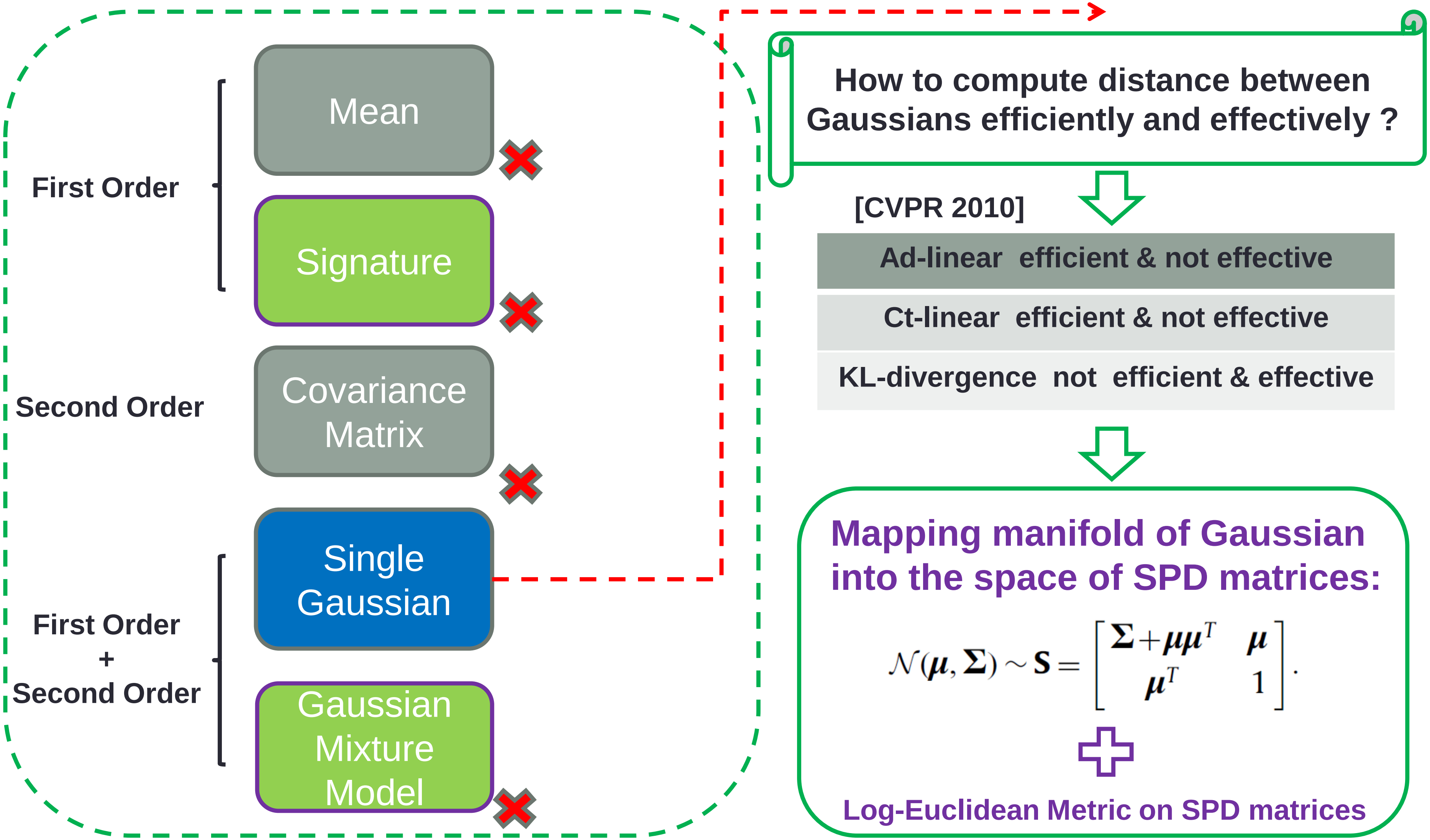


Gaussian

Outline

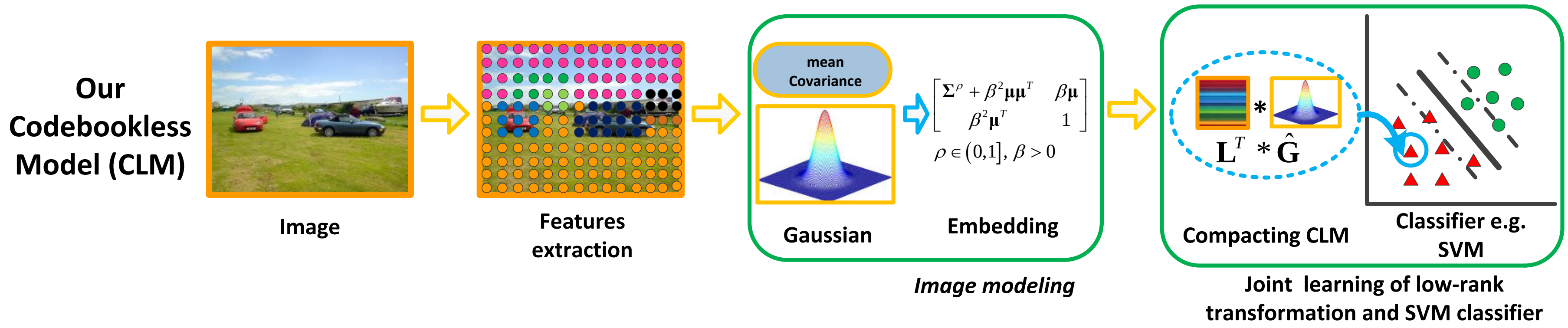
- ***Modeling Methods in Image Classification***
- ***Towards Effective Codebook-free Model***
- ***Robust Approximate Infinite Dimensional Gaussian***
- ***Future Work and Conclusion***

Metric between Gaussians



Peihua Li, Qilong Wang, Lei Zhang: A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models. ICCV, 2013.

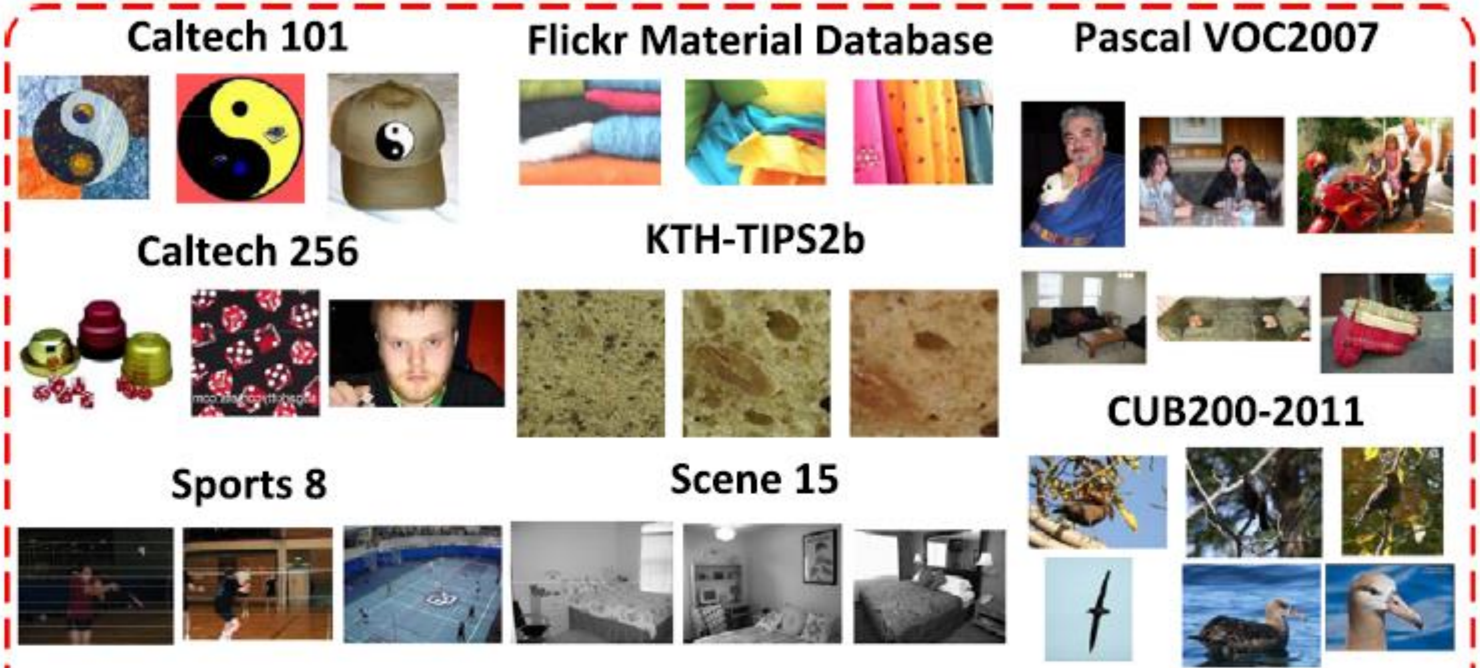
Pipeline of Proposed Method



1. Local (hand-crafted) features extraction.
2. Computing Gaussian and matching them with Embedding
3. Compacting CLM

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification. Pattern Recognition, 2016 (in press).

Comparison with the FV [IJCV13]



	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.

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification. Pattern Recognition, 2016 (in press).

Effect of Local Features

	Caltech101	Caltech256	VOC2007	CUB200-2011	FMD	KTH-TIPS-2b	Scene15	Sports8
FV+ SIFT	80.87+0.3	47.47+0.1	61.8	25.8	58.37+1.0	69.37+1.0	88.17+0.2	91.37+1.3
FV+ eSIFT	83.77+0.3	50.17+0.3	60.8	27.3	58.97+1.7	71.37+3.1	89.47+0.2	90.47+1.2
CLM + SIFT	84.97+0.1	48.97+0.2	55.8	18.6	51.67+1.2	71.87+3.1	88.17+0.4	88.87+1.0
CLM + eSIFT	86.37+0.3	53.67+0.2	60.4	28.1	57.77+1.6	75.27+2.6	89.47+0.4	91.57+1.2
CLM + L^2ECM	82.57+0.3	48.67+0.3	56.6	19.1	62.47+1.5	72.27+3.3	88.37+0.6	88.37+1.3
CLM + eL^2ECM	84.77+0.2	53.27+0.1	61.7	28.6	64.27+1.0	73.67+2.6	89.27+0.5	90.77+0.7

Peihua Li, Qilong Wang, Local log-Euclidean covariance matrix (L^2ECM) for image representation and its applications, in ECCV, 2012.
Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification. Pattern Recognition, 2016 (in press).

Comparison with counterparts

	Scene15	Sports8
GG (ad-linear) [CVPR2010]	79.8	80.2
GG (ct-linear) [CVPR2010]	82.3	82.9
GG (KL-kernel) [CVPR2010]	86.1	84.4
CLM (SIFT)	88.1	88.8

Metric between Gaussian models is very important.

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification. Pattern Recognition, 2016 (in press).

Some key findings

- ***Our work has clearly shown that single Gaussian is a very competitive alternative to the mainstream BoW model.***

- ***Comparison with BoW model, our method is more efficient with no requirement of dictionary. Meanwhile, it avoid aforementioned limitations of BoW model.***

- ***Our method is more suit for texture or material images.***

- ***More powerful local descriptors can bring more improvement for our method than BoF model.***

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification. Pattern Recognition, 2016 (in press).

Outline

- *Modeling Methods in Image Classification*
- *Towards Effective Codebook-free Model*
- ***Robust Approximate Infinite Dimensional Gaussian***
- *Future Work and Conclusion*

More Powerful Local Features

- Features from deep Convolutional Neural Network.
  - Fully-connected layer
    - MOP-CNN [ECCV 2014], SCFVC [NIPS2014], …
  - Convolutional layer
    - SPP-Net [ECCV 2014], FV-CNN [CVPR2015], …

- Infinite dimensional descriptors can provide richer and more discriminative information than their low dimensional counterparts.
  - Mapping local features into (approximated) RKHS
    - [CVPR2014], [NIPS2014], [ICASSP2015]

Approximate Infinite Dimensional Gaussian

Goal:

Computing infinite dimensional Gaussian with the features from deep Convolutional Neural Network.

Approximate Infinite Dimensional Gaussian

Goal: Computing infinite dimensional Gaussian with the features from deep Convolutional Neural Network.

Methods	Descriptor	Kernels or mappings	Estimator	Metric	Linear SVM ?
Zhou *et al.* [53]	Gaussian	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Probabilistic distances in $\mathcal{H}$	No
Harandi *et al.* [23]	Covariance	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Bregman Divergences in $\mathcal{H}$	No
Log-HS [20]	Covariance	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Log-Hilbert-Schmidt metric	No
Faraki *et al.* [17]	Covariance	{Random Fourier transform, Nyström method} for RBF kernel	Ledoit-Wolf estimator	Log-Euclidean metric	Yes
RAID-G (Ours)	Gaussian	Explicit feature maps of {Hellinger's kernel, $\mathcal{X}^2$ kernel}	Regularized MLE with von Neumann divergence	Gaussian Embedding and vectorization	Yes

Table 1. Comparison of different infinite dimensional image descriptors.

Approximate Infinite Dimensional Gaussian

Goal: Computing infinite dimensional Gaussian with the features from deep Convolutional Neural Network.

Methods	Descriptor	Kernels or mappings	Estimator	Metric	Linear SVM ?
Zhou *et al.* [53]	Gaussian	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Probabilistic distances in $\mathcal{H}$	No
Harandi *et al.* [23]	Covariance	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Bregman Divergences in $\mathcal{H}$	No
Log-HS [20]	Covariance	RBF kernel (no explicit mapping)	Ledoit-Wolf estimator	Log-Hilbert-Schmidt metric	No
Faraki *et al.* [17]	Covariance	{Random Fourier transform, Nyström method} for RBF kernel	Ledoit-Wolf estimator	Log-Euclidean metric	Yes
RAID-G (Ours)	Gaussian	Explicit feature maps of {Hellinger's kernel, $\mathcal{X}^2$ kernel}	Regularized MLE with von Neumann divergence	Gaussian Embedding and vectorization	Yes

Table 1. Comparison of different infinite dimensional image descriptors.

Our solution:

Two explicit feature mappings:

(1)

$$\phi_{Hel}(\mathbf{x}_k) = \sqrt{\mathbf{x}_k}.$$

(2)

$$\widehat{\phi}_{Chi}(\mathbf{x}_k) = \sqrt{\mathbf{x}_k}\left[\sqrt{L}, \sqrt{2L\mathrm{sech}(L\pi)}\cos(L\log(\mathbf{x}_k)) \right.$$
$$\left.\sqrt{2L\mathrm{sech}(L\pi)}\sin(L\log(\mathbf{x}_k))\right]^T, \quad ($$

Robust Estimation of Approximate Infinite Dimensional Gaussian

Problem:

We face to estimation of covariance in high dimensional problems with a small number of samples. It is well known that conventional Maximum Likelihood Estimation (MLE) is not robust to this condition.

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Robust Estimation of Approximate Infinite Dimensional Gaussian

Problem: **We face to estimation of covariance in high dimensional problems with a small number of samples. It is well known that conventional Maximum Likelihood Estimation (MLE) is not robust to this condition.**

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

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

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

⇩

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

Classical MLE

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Robust Estimation of Approximate Infinite Dimensional Gaussian

Problem: **We face to estimation of covariance in high dimensional problems with a small number of samples. It is well known that conventional Maximum Likelihood Estimation (MLE) is not robust to this condition.**

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

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

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

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

Classical MLE

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

where $D_{\mathrm{vN}}(\mathbf{A},\mathbf{B}) = \mathrm{tr}(\mathbf{A}(\log(\mathbf{A})-\log(\mathbf{B}))-\mathbf{A}+\mathbf{B})$

is the von Neumann divergence between matrices.

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

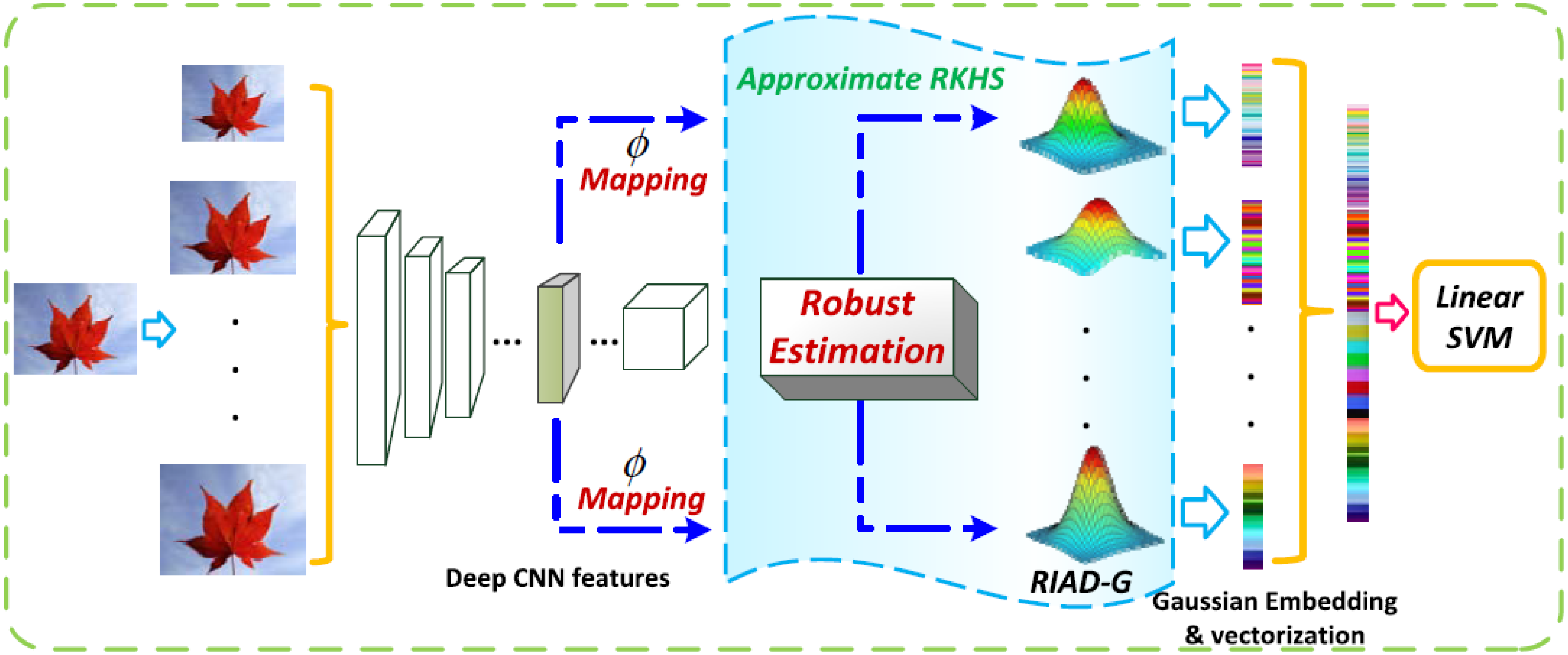
$$\lambda_k = \sqrt{\Big(\frac{1-\alpha}{2\alpha}\Big)^2 + \frac{\delta_k}{\alpha}} - \frac{1-\alpha}{2\alpha}$$

$$\widehat{\mathbf{S}} = \widehat{\mathbf{U}}\mathrm{diag}(\delta_k)\widehat{\mathbf{U}}^T$$

vN-MLE

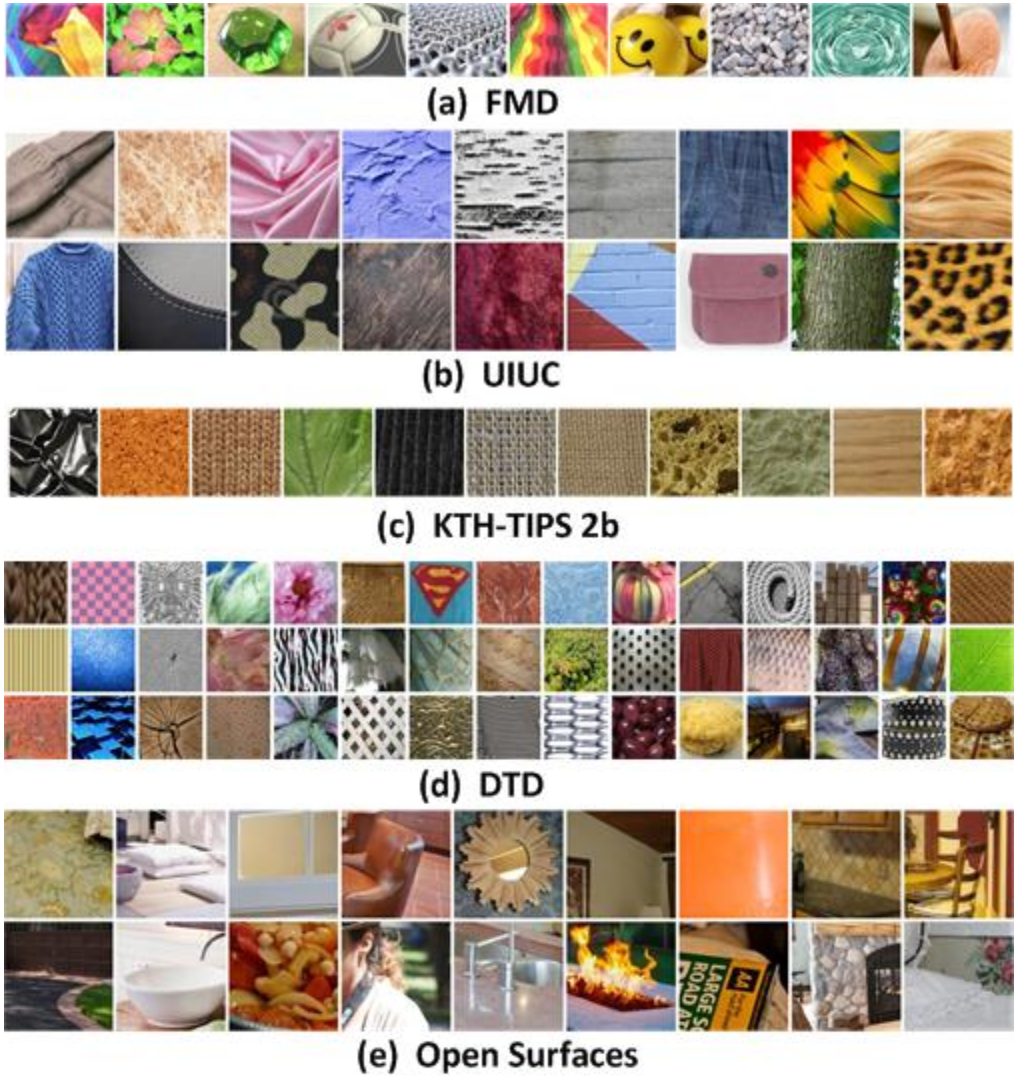
Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Connection with Other Infinite Dimensional Models



Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Material Recognition



(a) FMD
(b) UIUC
(c) KTH-TIPS 2b
(d) DTD
(e) Open Surfaces

Results on Material Recognition

The accuracy (%) of various methods on five material benchmarks. $*$: The score level fusion is used to combine FC and FV-CNN.

VGG-VD-16 without fine-tuning

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]	-

- **Gaussian descriptors > covariance descriptors.**
- **The proposed vN-MLE estimator can achieve big performance improvements.**
- **Gaussian descriptors constructed in RKHS > those constructed in the original space.**
- **RAID-G outperforms FV-CNN and achieves state-of-the-art performances.**

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Robust Covariance Estimation

Comparison with various robust estimators on FMD and UIUC material databases.

Methods	FMD	UIUC Material
Gau-CNN (LW)	81.3 ± 1.4	81.7 ± 2.9
Gau-CNN (Stein)	81.9 ± 0.7	82.2 ± 1.8
Gau-CNN (MMSE)	81.2 ± 1.2	80.9 ± 1.9
Gau-CNN (EL-SP)	81.5 ± 1.6	82.0 ± 2.3
RoG-CNN (vN-MLE)	$\mathbf{83.6 \pm 1.6}$	$\mathbf{84.5 \pm 1.8}$
Gau-CNN-Chi (LW)	83.1 ± 0.9	81.6 ± 4.1
Gau-CNN-Chi (Stein)	83.2 ± 0.8	83.6 ± 3.0
Gau-CNN-Chi (MMSE)	83.1 ± 0.8	82.0 ± 4.3
Gau-CNN-Chi (EL-SP)	83.2 ± 1.1	82.1 ± 3.1
RAID-G-CNN-Chi (vN-MLE)	$\mathbf{84.9 \pm 1.4}$	$\mathbf{86.3 \pm 2.9}$

The vN-MLE is superior to the competing methods in the very high dimensional setting.

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Explicit Feature Mappings

Effects of various feature mappings on FMD and UIUC material database.

Methods	FMD	UIUC Material
RAID-G-CNN-rFt (1x)	79.7 ± 1.6	80.6 ± 2.2
RAID-G-CNN-rFt (3x)	80.6 ± 2.3	81.8 ± 2.7
RAID-G-CNN-Nyström (1x)	82.2 ± 2.2	83.3 ± 3.1
RAID-G-CNN-Nyström (3x)	82.8 ± 1.9	84.0 ± 2.7
RAID-G-CNN-Hel	$\mathbf{84.4 \pm 1.3}$	$\mathbf{85.7 \pm 2.1}$
RAID-G-CNN-Chi	$\mathbf{84.9 \pm 1.4}$	$\mathbf{86.3 \pm 2.9}$
CDL_{rFt} [17]	-	47.4 ± 3.1
$CDL_{Nyström}$ [17]	-	46.3 ± 2.6

The introduced feature mappings are not only efficient but effective in very high dimensional setting.

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Infinite dimensional descriptors

Methods	Accuracy (in %)
RAID-G-Hel (23D Handcrafted features)	78.8 ± 4.8
RAID-G-Chi (23D Handcrafted features)	78.2 ± 4.7
RAID-G-CNN-Hel	$\mathbf{89.0 \pm 5.4}$
RAID-G-CNN-Chi	$\mathbf{89.3 \pm 4.5}$
Log-E RBF (baseline) (23D Handcrafted features)	74.1 ± 7.4
Harandi *et al.* [23] (23D Handcrafted features)	80.1 ± 4.6
Log-HS [20] (23D Handcrafted features)	81.9 ± 3.3

- **When hand-crafted features are used, the methods in [23, 20] are slightly better than RAID-G.**
- **When employing high dimensional deep CNN features, RAID-G achieves more than 7% improvements over infinite dimensional covariance descriptors [23, 20], where CNN features cannot be used due to unaffordable cost.**

Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Application to other tasks

VGG-VD-16 without fine-tuning

	CUB200-2011	Indoor67	SUN397
RIAD-G	82.1	82.8	67.1

排行榜

第 2 赛季排行榜 | 第 1 赛季排行榜

排名	参赛者	所在组织	评分
1	VIPL-3L	中国科学院	0.4929
2	DLUT_VLG	大连理工大学	0.4916
3	Choroi	复旦大学	0.4554
4	Hitsz_BCC	哈尔滨工业大学	0.4195
5	ToSsBoY	复旦大学	0.3801
6	KNIGHT-BUPT	北京邮电大学	0.3720
7	tjucs_lemon	天津大学	0.3579
8	nus_next	哈尔滨工业大学	0.3535

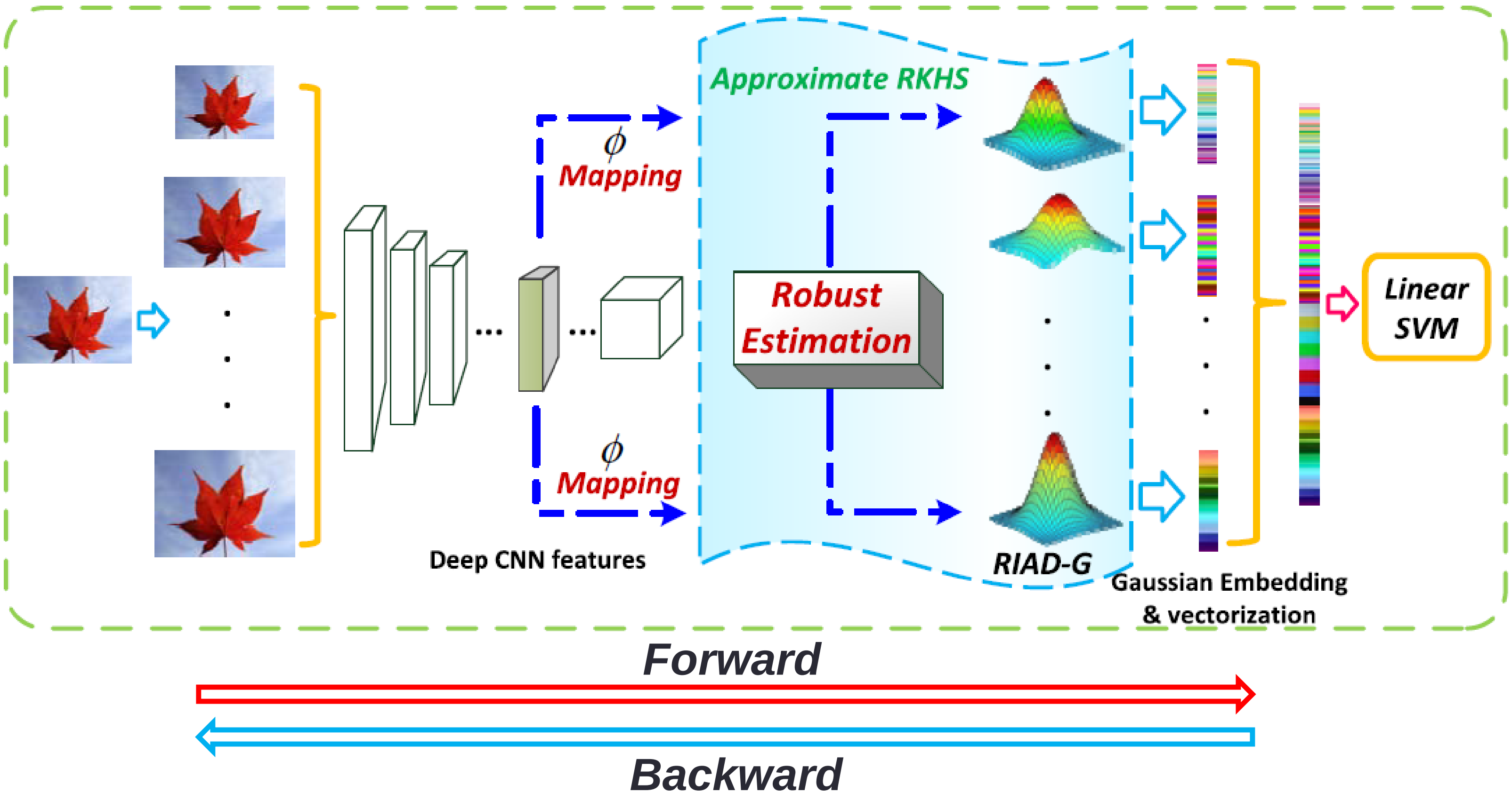
Qilong Wang, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016

Outline

- ***Modeling Methods in Image Classification***
- ***Towards Effective Codebook-free Model***
- ***Robust Approximate Infinite Dimensional Gaussian***
- ***Future Work and Conclusion***

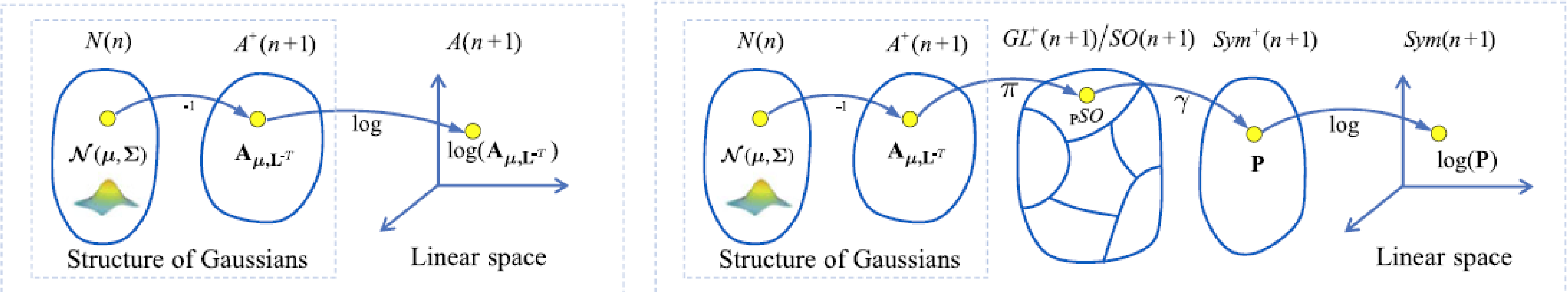
Future work

An end-to-end learning architecture



Future work

The better usage and understanding of manifold structure of Gaussian



(b) DE-LogE (c) IE-LogE

We show, for the first time to our knowledge, that the space of Gaussians can be equipped with a Lie group structure by defining a multiplication operation on this manifold.

Peihua Li, Qilong Wang, Hui Zeng and Lei Zhang, Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016 (in press).

Summary

- ***The codebook-free single Gaussian is a very competitive image model for classification, and is more sensitive to powerful local features.***

- ***Follow the similar pipeline, we proposed RIAD-G, a reinforced codebook-free single Gaussian model, with considering robust estimation of very high dimensional covariance matrix.***

- ***Now, we are trying to conduct a end-to-end learning architecture for RIAD-G to further improvement.***

- ***The better usage of manifold structure of Gaussian and more general model are mainly directions in our future work.***

Related References

- **Qilong Wang**, Peihua Li, Wangmeng Zuo, and Lei Zhang. RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition, In CVPR, 2016 (accepted).
- Peihua Li, **Qilong Wang**, Hui Zeng and Lei Zhang, Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016 (in press).
- **Qilong Wang**, Peihua Li, Wangmeng Zuo, and Lei Zhang. Towards effective codebookless model for image classification, Pattern Recognition, 2016 (in press).
- Peihua Li, **Qilong Wang**, Local log-Euclidean covariance matrix (L^2ECM) for image representation and its applications, in ECCV, 2012.
- Peihua Li, **Qilong Wang**, Lei Zhang: A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models, in ICCV, 2013.

端午节快乐

HAPPY DRAGON BOAT FESTIVAL

THANKS & QUESTIONS?

The codes can be downloaded at http://ice.dlut.edu.cn/PeihuaLi/