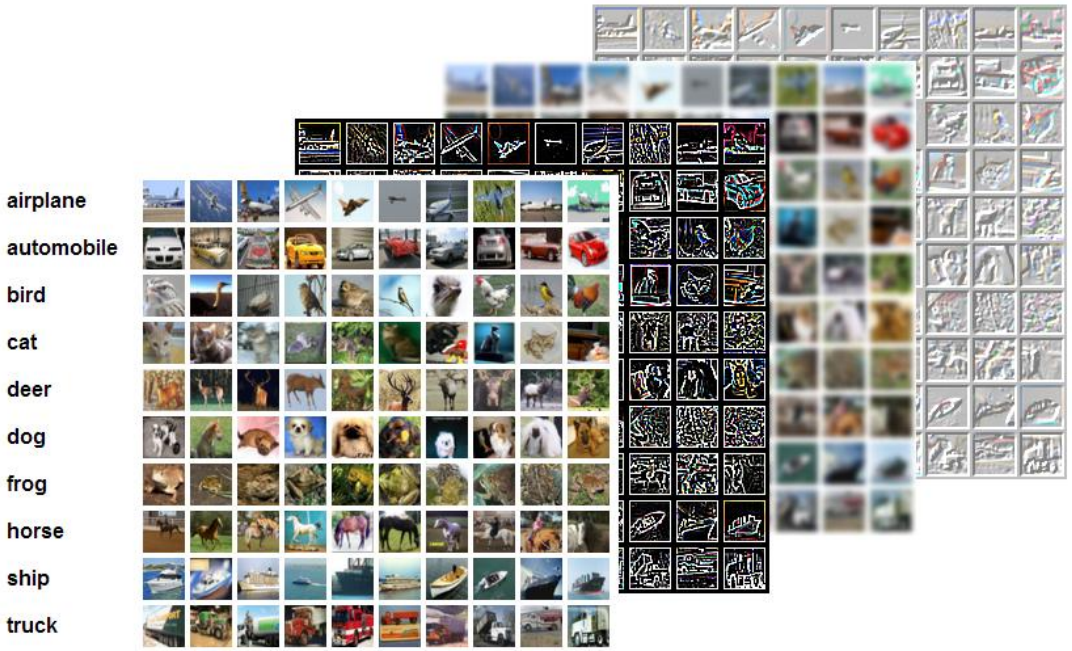
Université de Liège



Classification générique d'images : Approches aléatoires et convolutionnelles

Begon Jean-Michel

Mémoire de fin d'études réalisé en vue de l'obtention du grade de Master en sciences informatiques

UNIVERSITÉ DE LIÈGE

Faculté des sciences appliquées
Année académique 2013-2014

Plan

- Introduction
- La méthode RandConv
- Les modes de classification
- Base de données d'évaluation
- Résultats
- Conclusion et perspectives

Introduction

- Classification générique d'images
  - Apprentissage automatique

Objet	X1	X2	X3	…	Xn	Classe
#1	45	10	11	…	255	Chien
#2	255	27	200	…	14	Bateau
…	…	…	…	…	…	…
#m	16	0	4	…	144	Avion

- Images (données structurées)
  - Transformer la collection d'objets en matrice d'apprentissage
    - Descripteur d'images

La méthode RandConv

- Objectif :
  - Combiner les avantages de la méthode Pixit
    - Simplicité
    - Temps de calcul (et d'optimisation)
  - à ceux des réseaux de neurones à convolution (ConvNets)
    - Exactitude

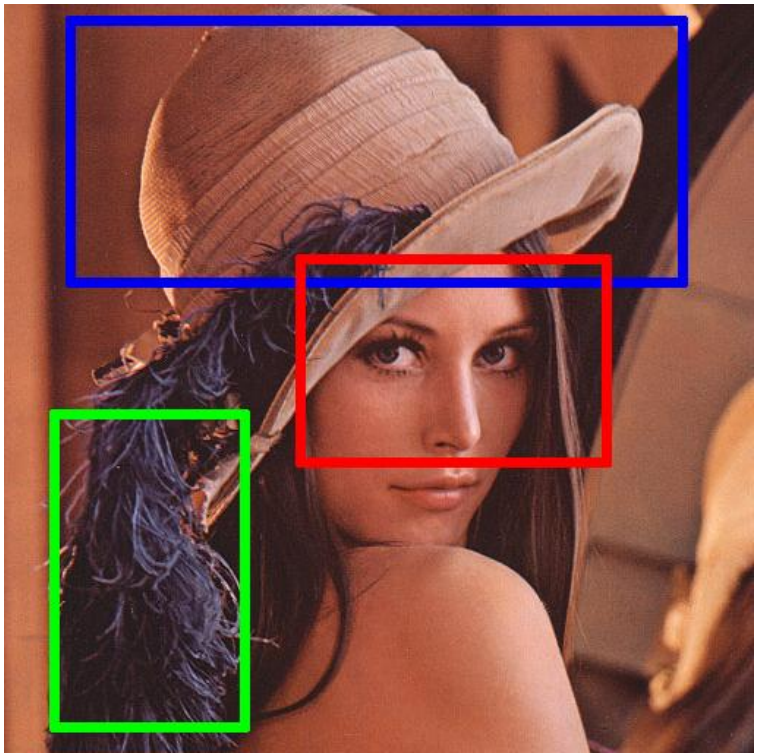
- Principe :
  - Application de filtres aléatoires
  - Échantillonnage spatial

  ConvNets

  - Extraction de sous-fenêtres et redimensionnement
  - Description par les pixels bruts

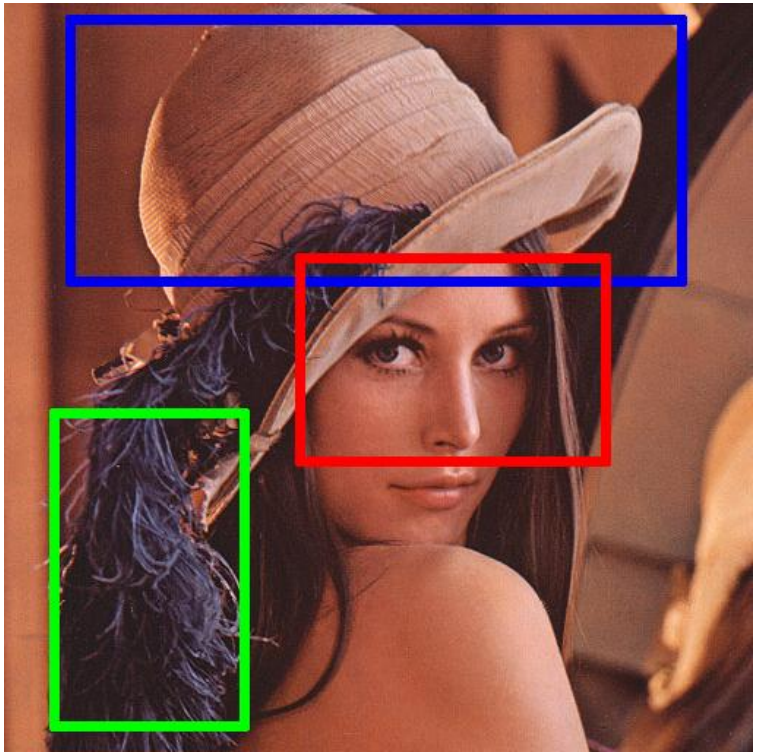
  Pixit

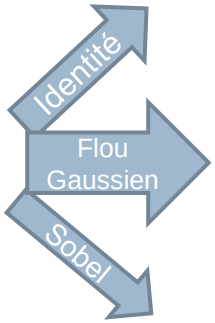
La méthode RandConv : Sélection des sous-fenêtres (Pixit)

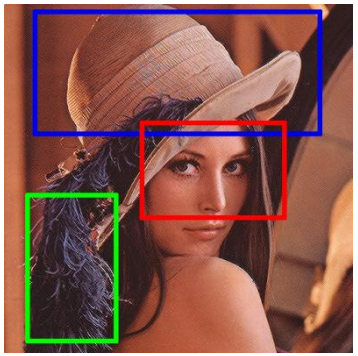


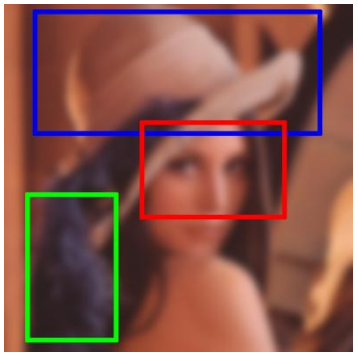
- Tirage aléatoire uniforme

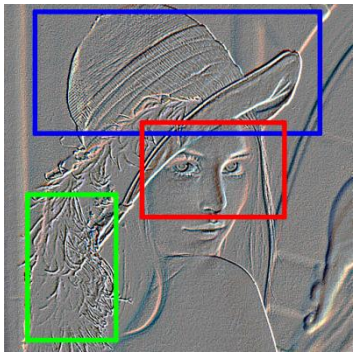
La méthode RandConv : Filtrage (ConvNets)



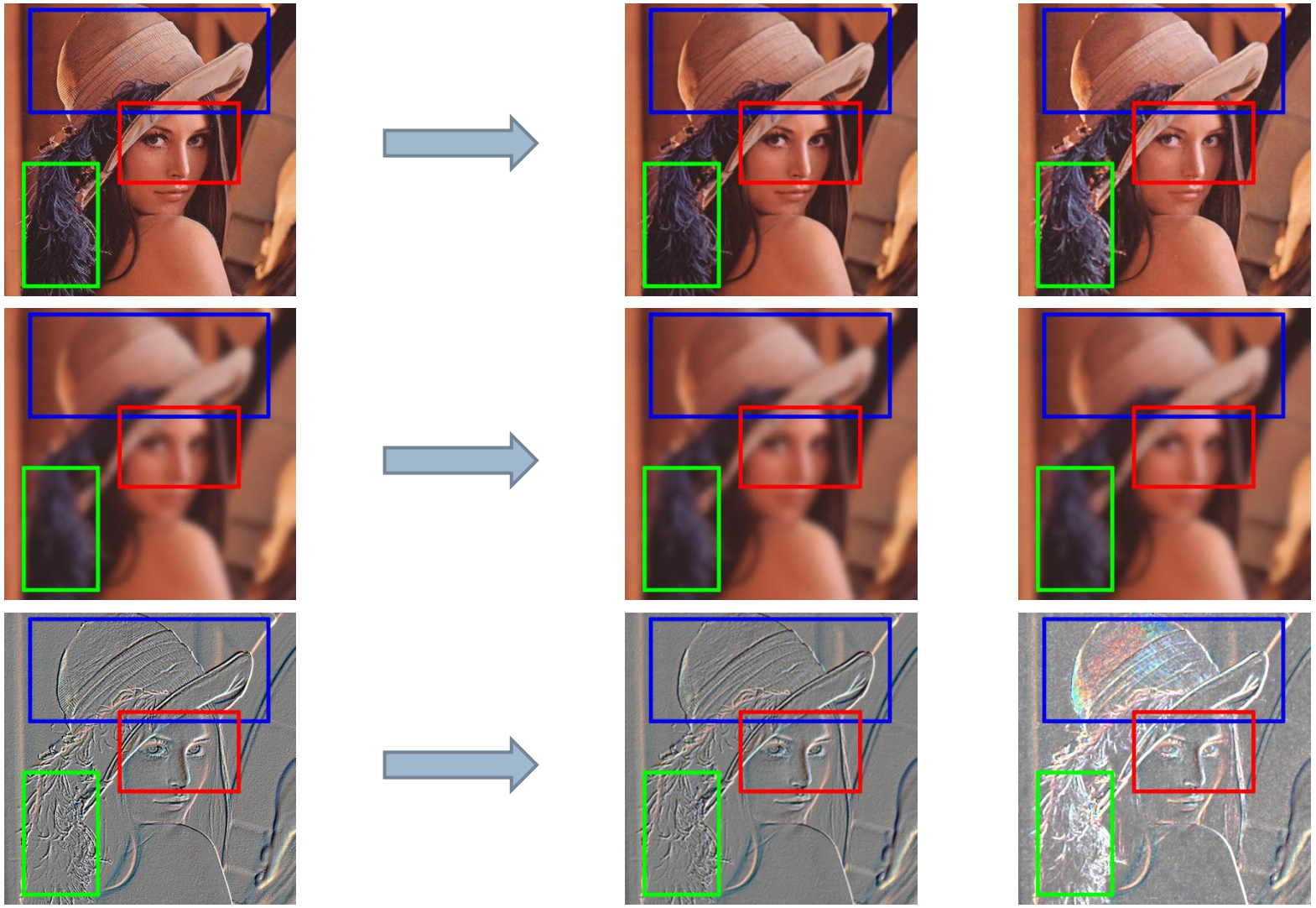








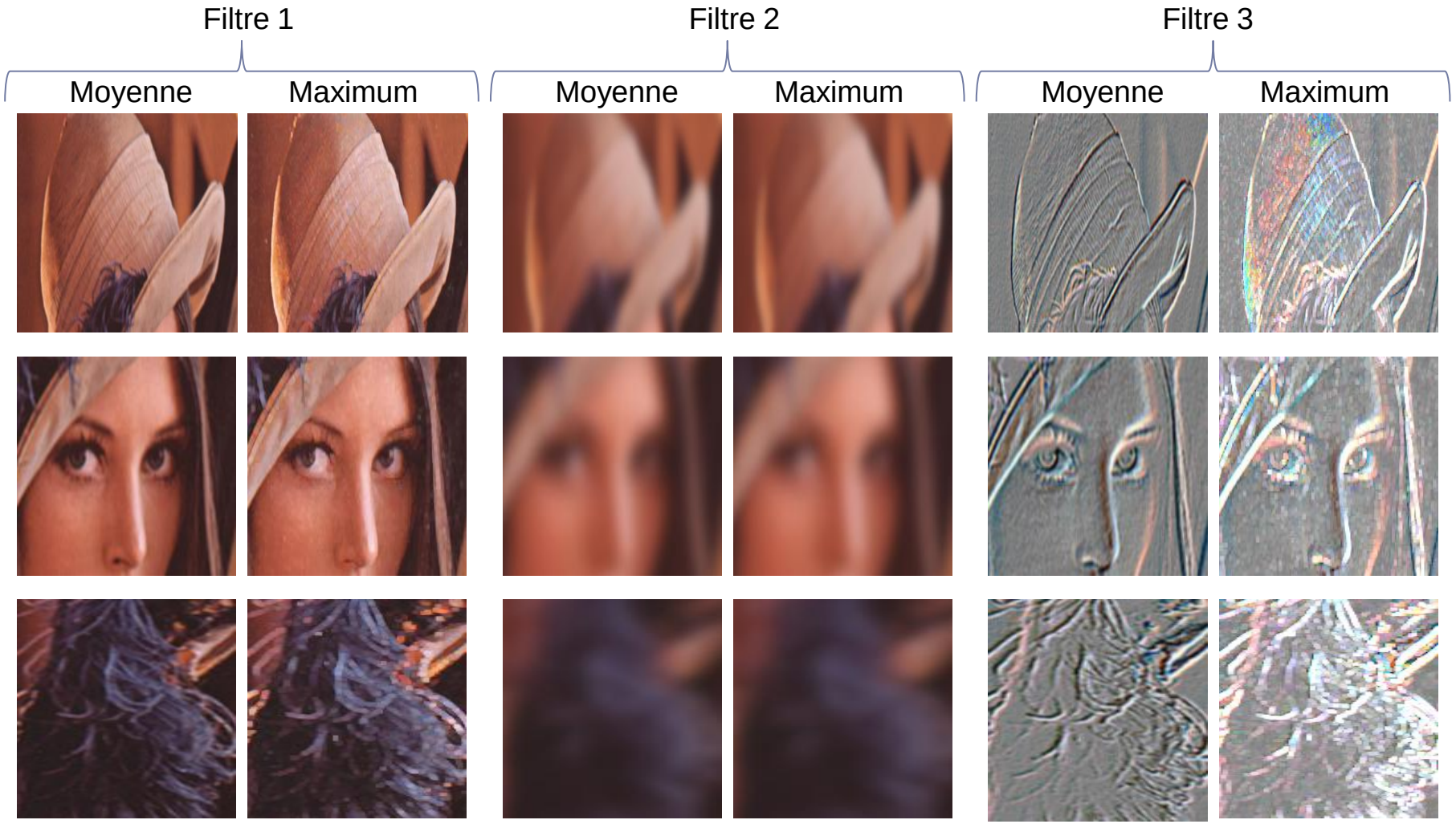
La méthode RandConv : Échantillonnage spatial (ConvNets)



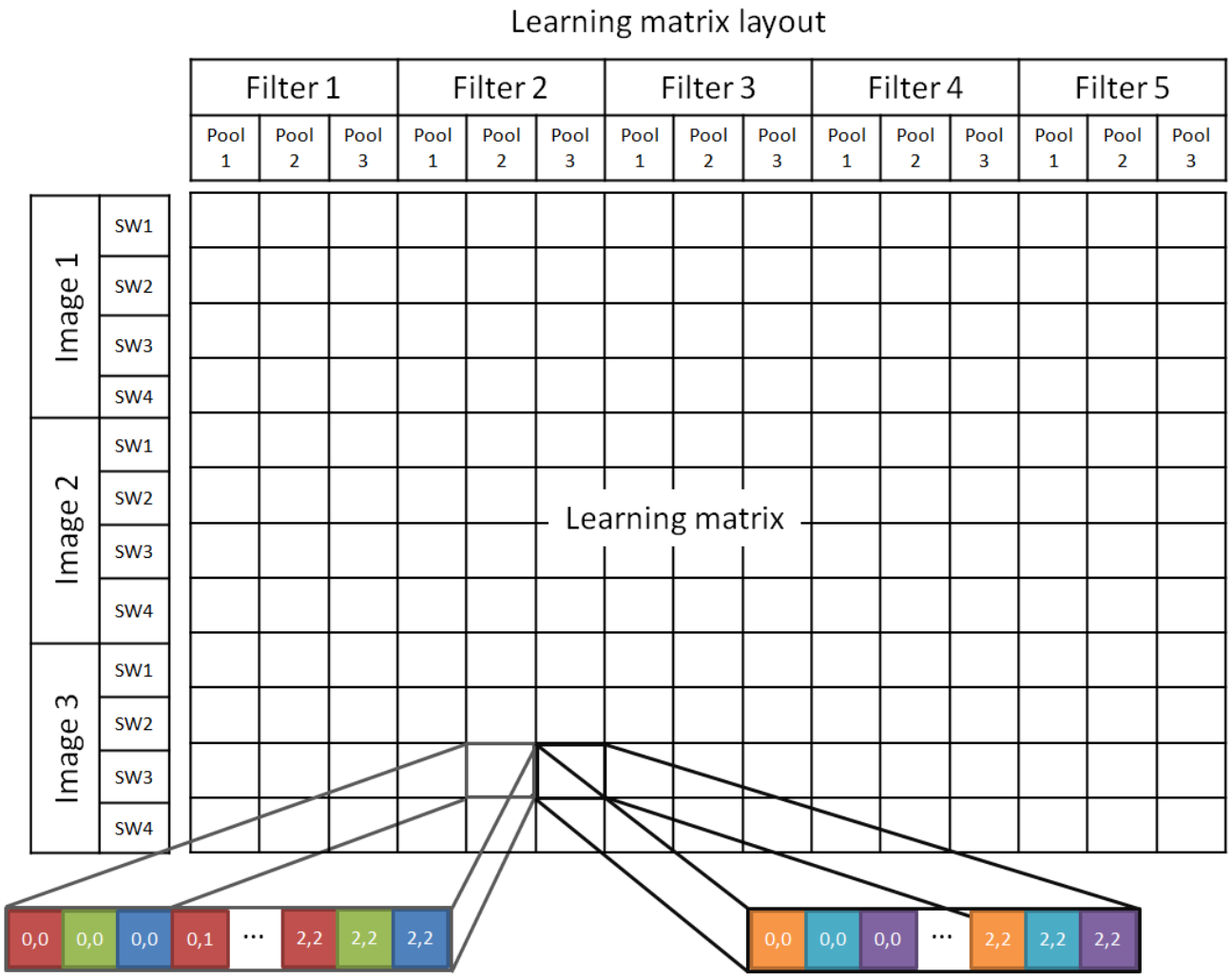
Moyenne

Maximum

La méthode RandConv : Extraction et redimensionnement des sous-fenêtres (Pixit)



La méthode RandConv : Construction de la matrice d'apprentissage

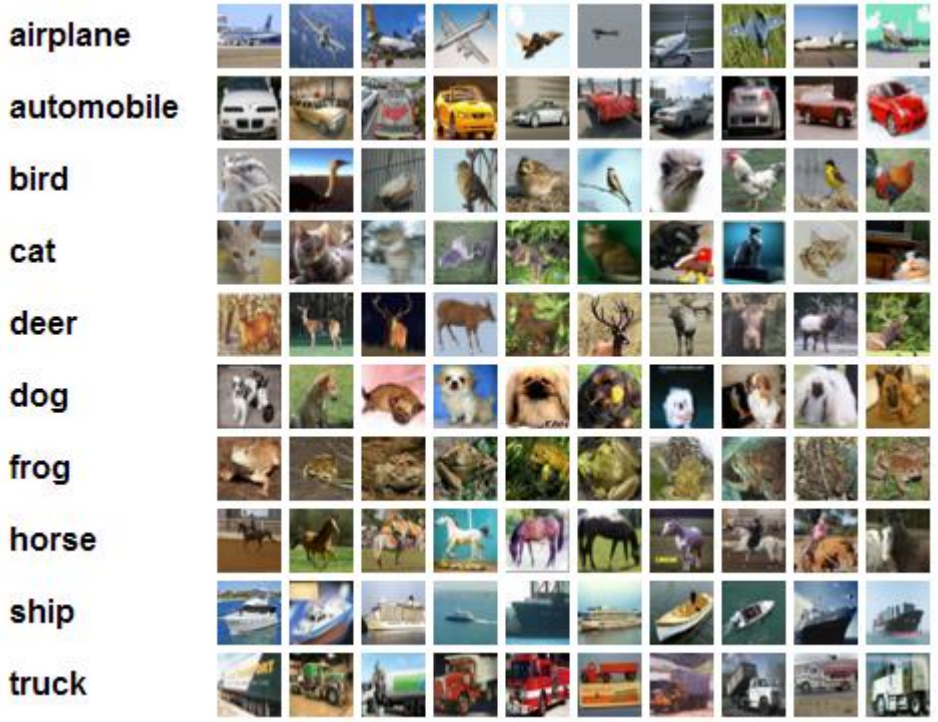


Les modes de classification

- Deux modes
  - ET-DIC (*ExtraTrees for Direct Image Classification*)
    - Soumission de la matrice d'apprentissage aux *ExtraTrees*
    - Agrégation des prédictions des sous-fenêtres d'une même image (comme Pixit)
  - ET-FL (*ExtraTrees for Feature Learning*)
    - Création d'un dictionnaire (*ExtraTrees*)
    - Histogramme de « mots visuels » sur base des sous-fenêtres d'une même image
    - Classification par une machine à vecteurs de support (SVM)

Base de données d'évaluation : CIFAR-10

- 50 000 images d'apprentissage
- 10 000 images de test
- 32x32 RGB
- 10 catégories
- Difficile pour Pixit
  - 53,67%
- Meilleur résultat :
  - 91.2% (ConvNets)



Résultats du mode ET-DIC

- Meilleurs résultats de Pixit (ET-DIC)
  - 53.67% (20 sous-fenêtres, 10 arbres, n_min = 10)
  - 49.77% (10 sous-fenêtres, 30 arbres complètement développés)
- RandConv
  - 51.51% (10 sous-fenêtres, 30 arbres complètement développés, 100 filtres aléatoires, fenêtre glissante moyennante 3x3)
  - Faible variabilité de l'exactitude
  - Grande stabilité de l'importance des filtres

Peu important	Important
Filtres (nombre et type)	Paramètres des arbres Nombre de sous-fenêtres Échantillonnage spatial (**fonction**, échelle)

Résultats du mode ET-DIC

- RandConv
  - Meilleur résultat : 63,30% (~24h de calculs)
    - Optimisation des paramètres
  - Présélection des filtres
    - Inutile
  - Combinaison de forêts
    - Gain lié au nombre d'arbres

Résultats du mode ET-FL

- Pixit : 58.79%
  - 750 *Totally randomized trees*, n_min = 500, fenêtre glissante moyennante 3x3
- RandConv :
  - 63.55%
    - Filtres prédéfinis, 750 *Totally randomized trees*, n_min = 500, fenêtre glissante moyennante 3x3
  - 55,07 – 59.37%
    - Filtres aléatoires, 750 *Totally randomized trees*, n_min = 500, fenêtre glissante moyennante 3x3
- Rôle important des filtres
  - Disparition de la sélection des « bons » filtres
  - → Chute d'exactitude
  - → Grande variabilité
- Exactitude – taille du dictionnaire
  - Tendance linéaire

Résultats du mode ET-FL

- Améliorations
  - Présélection des filtres importants
    - 55,07 – 59.37 → 62.11%
  - Combinaison de plusieurs ensembles de filtres (63.98%, 63.69%, 61.61%)
    - Méthode d'ensemble : 67.10%
    - Construction de plusieurs forêts et agrégation
      - Dictionnaire de même ordre de taille : 64.59%
      - Dictionnaire complet : 68.99%

Résultats du mode ET-FL

- Meilleur résultat
  - Présélection de 60 filtres aléatoires parmi 500 avec fenêtres glissantes maximisantes 3x3 et 7x7
    - 70.74%
  - Filtres prédéfinis avec fenêtres maximum glissantes 3x3 et 7x7
    - 70.62%, 70.57%, 70.87%, 70.70%, 70.97%
  - Combinaison « *leave-one-out* » pour la création de dictionnaires complets
    - 73.97% - 74.27%
  - Méthode d'ensemble sur les résultats
    - 74.40%

Conclusion

- Objectif :
  - Combiner les avantages de la méthode Pixit à ceux des ConvNets

Sous-objectifs	ET-DIC	ET-FL
Simplicité	Oui (par rapport aux ConvNets)	
Temps	Pixit << RandConv < ConvNets	
Exactitude	Pixit < RC-DIC << ConvNets	Pixit < RC-FL < ConvNets

Perspectives

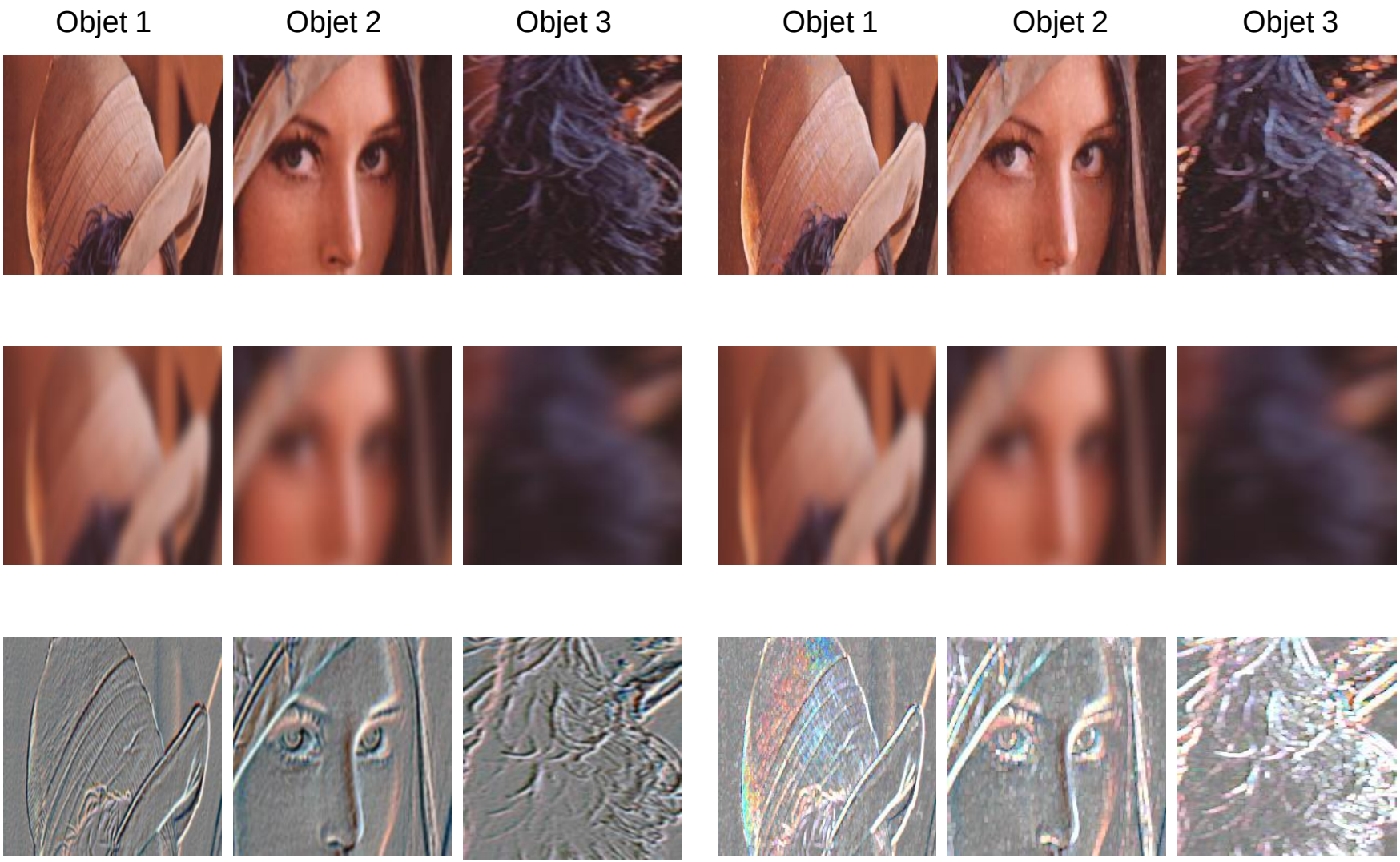
- Approfondir le mode ET-FL
- Confirmer les observations
- Introduire de la non-linéarité
- Nouveaux générateurs de filtres linéaires basés sur le domaine fréquentiel

- D’autres approches sont envisageables pour combiner les deux méthodes

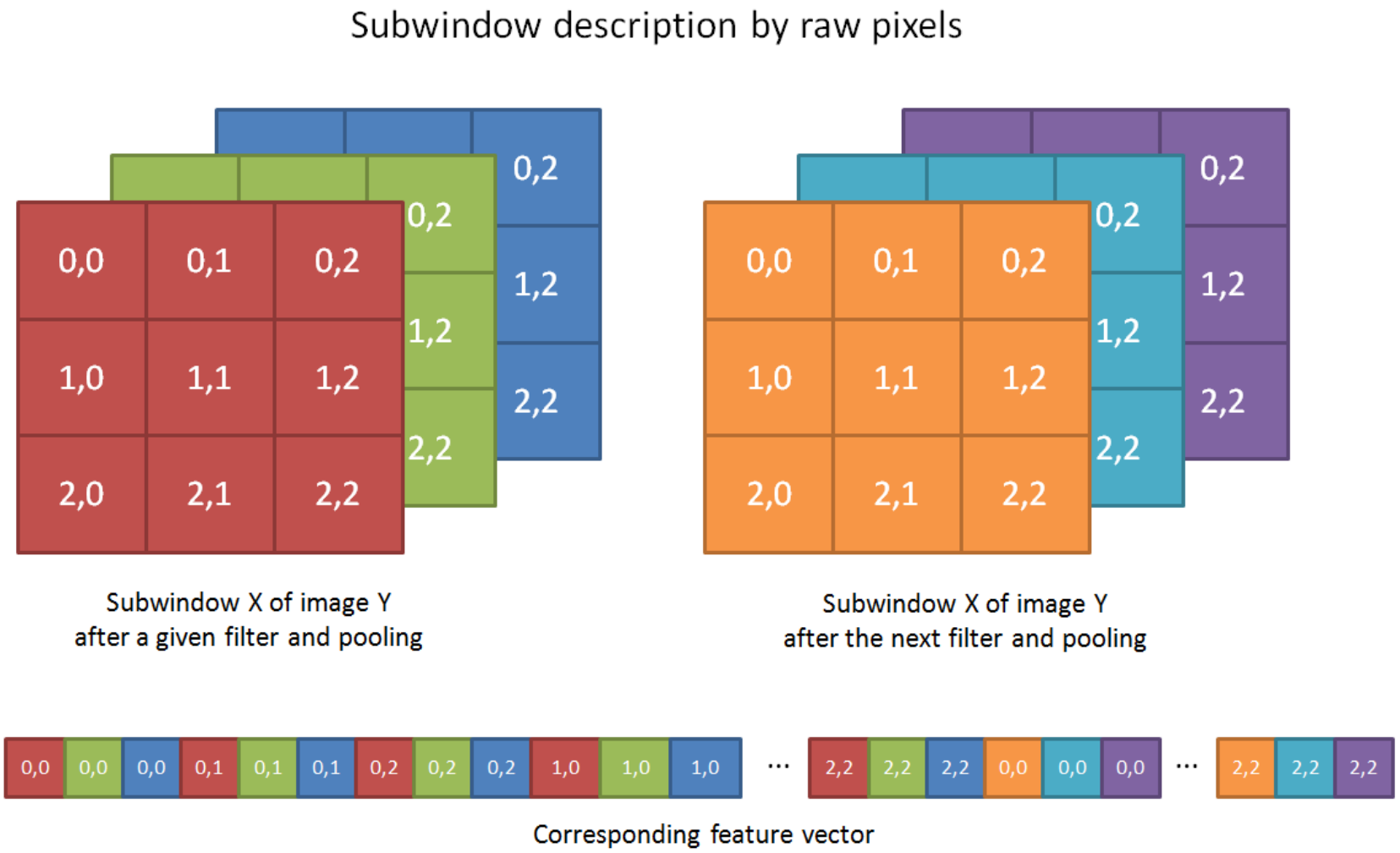
Merci pour votre attention

Slides additionels:

La méthode RandConv : Extraction et redimensionnement des sous-fenêtres



La méthode RandConv : Construction de la matrice d'apprentissage (Pixit)

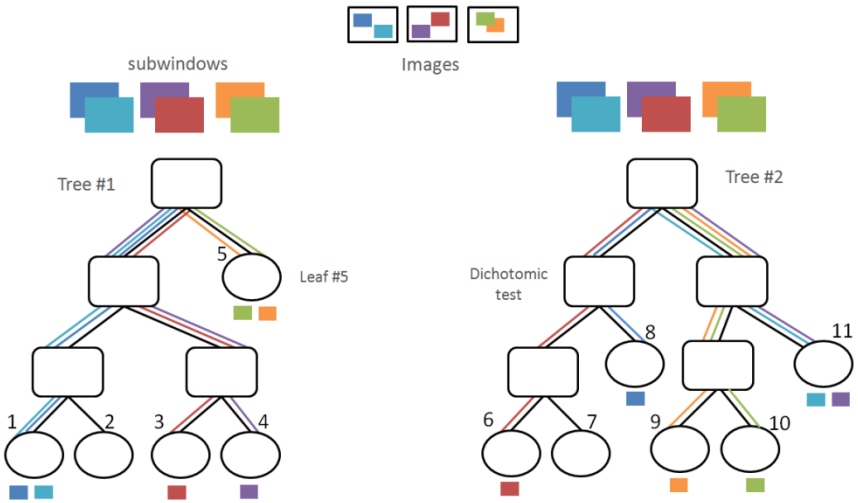


La méthode RandConv : remarques

- Génération des filtres
  - Prédéfinis
  - Centrés sur le filtre nul (perturbation)
  - Centrés sur le filtre identité
    - Perturbation
    - Distance
  - Stratifiés
- Échantillonnage spatial
  - Agrégation (voisinage sans chevauchement)
  - Fenêtre glissante (chevauchement)
  - Minimum, moyenne, maximum
- Limitation
  - Coût en mémoire
  - Coût des temps de calcul

Les modes de classification : ET-FL

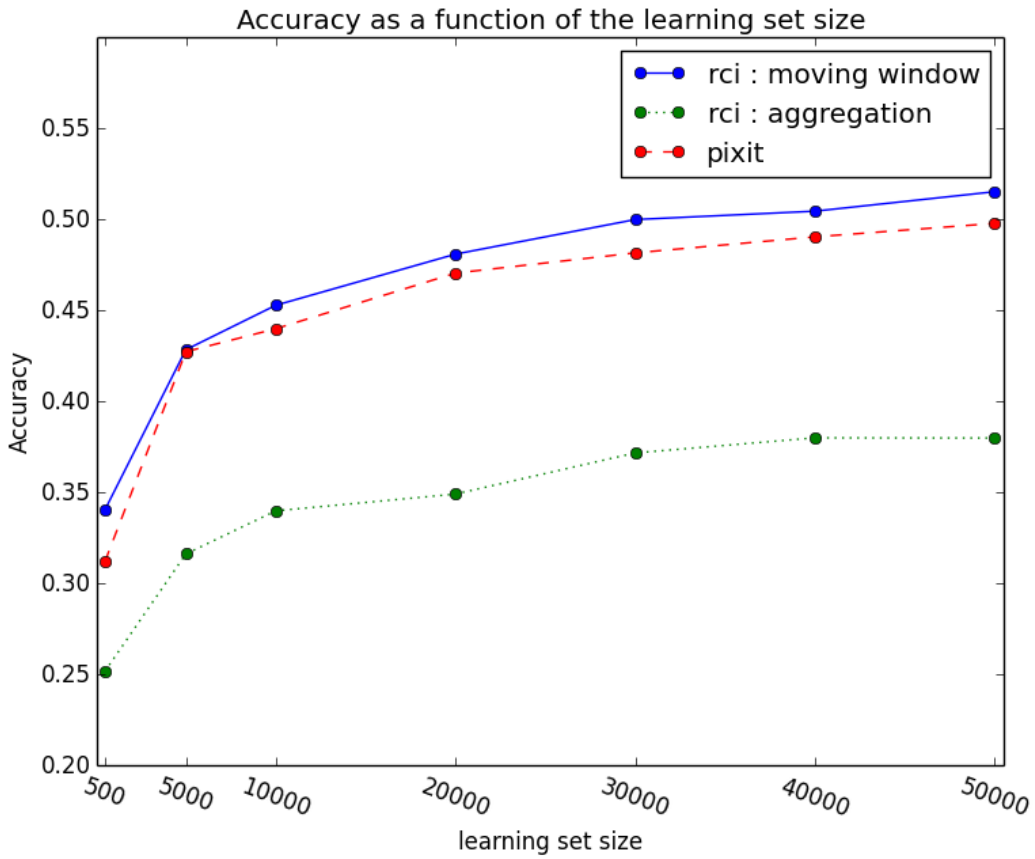
ET-FL : construction of the histogram of visual words



		Leaves										
		1	2	3	4	5	6	7	8	9	10	11
Image 1	Sw. 1	I							I			
	Sw. 2	I										I
	Final word	2	0	0	0	0	0	0	1	0	0	1
Image 2	Sw. 1			I			I					
	Sw. 2				I							I
	Final word	0	0	1	1	0	1	0	0	0	0	1
Image 3	Sw. 1					I					I	
	Sw. 2					I				I		
	Final word	0	0	0	0	2	0	0	0	1	1	0

Résultats du mode ET-DIC

- Influence de la taille de l'ensemble d'apprentissage
  - Agrégation : inefficace
  - Légère dominance de RandConv par rapport à Pixit



-30 arbres non-élagués
-10 sous-fenêtres
-100 filtres centrés sur le filtre nul et perturbés uniformément
-Fenêtre mobile moyennante 3x3

Résultats du mode ET-DIC

- Variabilité
  - Liée aux arbres

Statistique	Exactitude
Minimum	51.14%
Moyenne	51.36%
Maximum	51.56%
Écart-type	0.00148

    - Grande stabilité des importances de filtres
  - Liée à la matrice d'apprentissage (pour un générateur fixé)

Statistique	Exactitude
Minimum	50.93%
Moyenne	51.23%
Maximum	51.83%
Écart-type	0.00257

Résultats du mode ET-DIC

- Influence des paramètres des arbres
  - Nombre d'arbres
    - Accroissement monotone de l'exactitude
    - Stabilité de l'importance des filtres
  - Taille du sous espace aléatoire de variables (*k*)
    - Accroissement de l'exactitude avec la taille
    - Stabilité de l'importance des filtres
  - Profondeur des arbres (*n_min*)
    - Accroissement de l'exactitude avec la profondeur
    - Stabilité mitigée de l'importance des filtres
- Influence du nombre de sous-fenêtres
  - Accroissement de l'exactitude
  - Stabilité des filtres

N	Exact.
10	0.492
30 (déf.)	0.515
500	0.529

K	Exact.
1	0.447
277 (déf.)	0.515
10,000	0.528

N_min	Exact.
2 (déf.)	0.515
50	0.498
500	0.453

N	Exact.
5	0.48
10 (déf.)	0.515
18	0.521

Résultats du mode ET-DIC

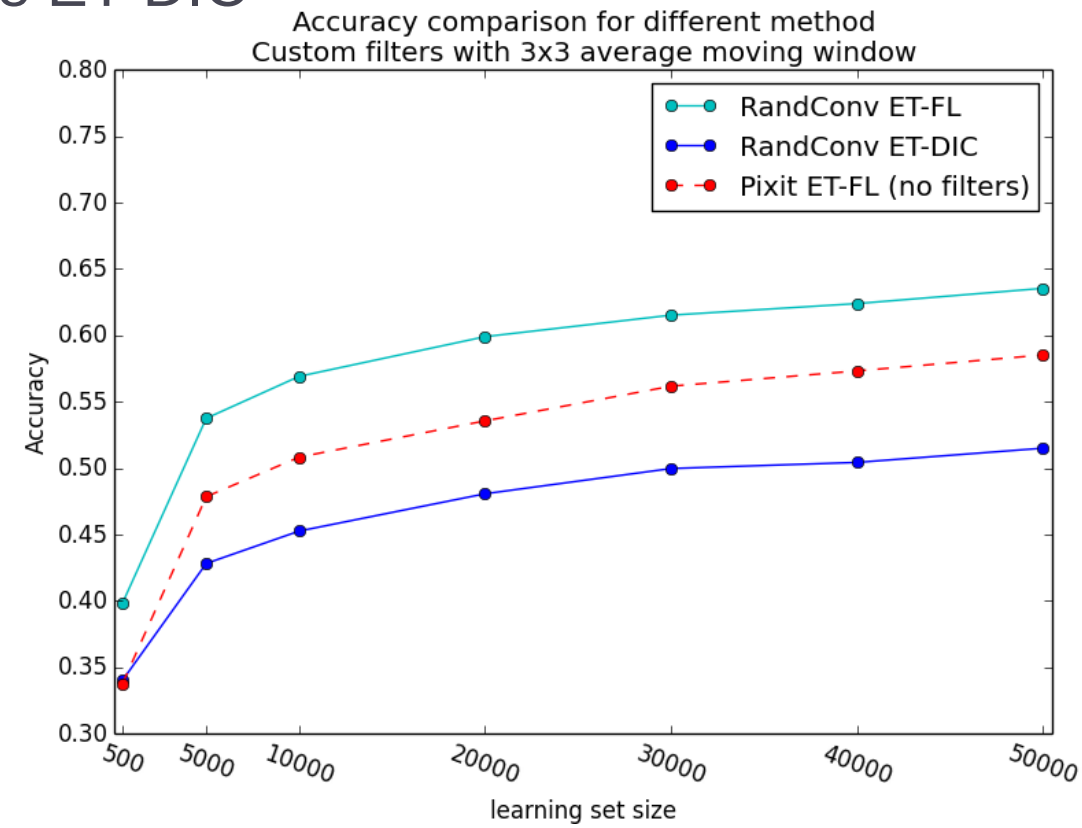
- Influence des paramètres convolutionels
  - Nombre de filtres
    - Influence faible sur l'exactitude
  - Taille des filtres
    - Préférence des petites tailles (3x3 – 9x9) pour CIFAR-10
  - Influence du générateur
    - Faible
    - Générateurs basés sur la distance : moins bons
  - Influence de l'échantillonnage spatial
    - Importante
    - Fonction maximum
    - Meilleur résultat : 57.41% fenêtre maximum glissante 3 à échelles (3x3, 5x5 et 7x7)
    - Importances des filtres varient avec la fonction mais peu avec l'échelle

Résultats du mode ET-DIC

- Borne maximum en optimisant les paramètres
  - 63,30% (~24h de calculs)
- Améliorations
  - Présélection des filtres importants
    - 51,37%
  - Création de plusieurs forêts et agrégation des résultats
    - Combinaisons des 3 générateurs de filtres différents (51.28%, 49.99% et 50.63%) :
      - 52.41% au final
        - A relativiser avec le nombre d'arbres

Résultats du mode ET-FL

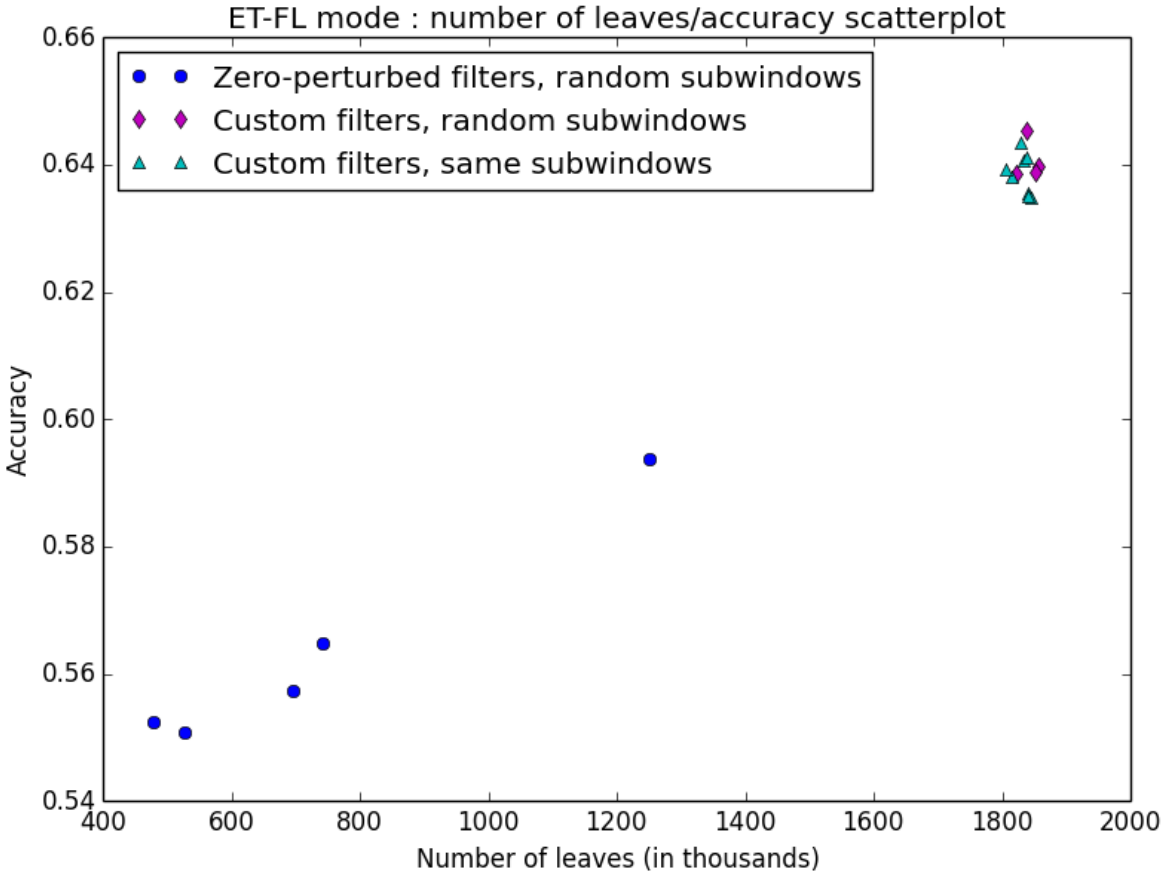
- Influence de la taille de l'ensemble d'apprentissage
  - Filtres prédéfinis
  - Meilleur que le mode ET-DIC
  - Domine le Pixit en ET-FL
  - 63.55%



-750 arbres
-n_min = 500
-20 sous-fenêtres
-38 filtres prédéfinis
-Fenêtre mobile moyennante 3x3

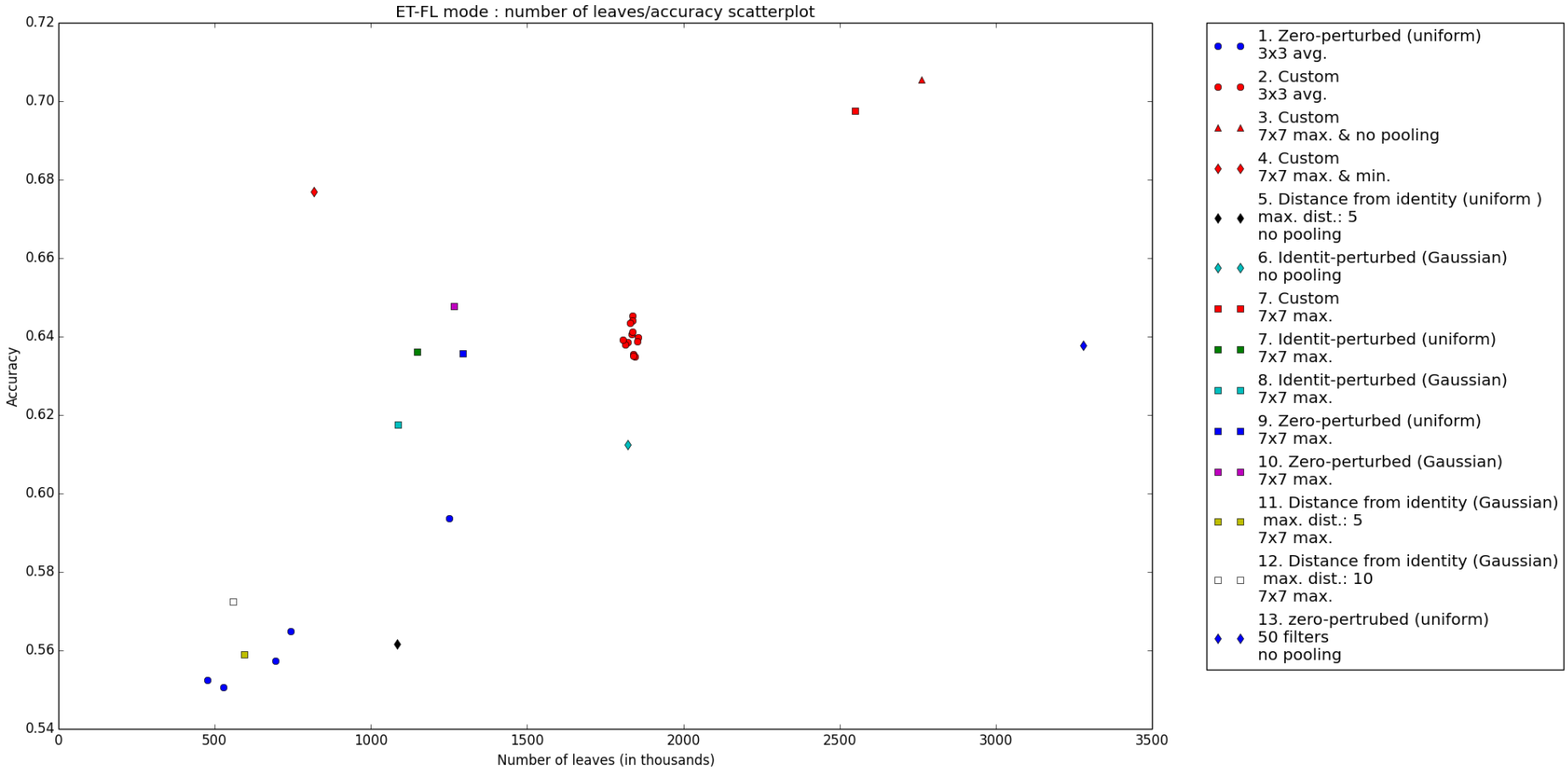
Résultats du mode ET-FL

- Variabilité
  - Plus importante qu'en ET-DIC
  - Faible influence des sous-fenêtres
  - Forte influence des filtres



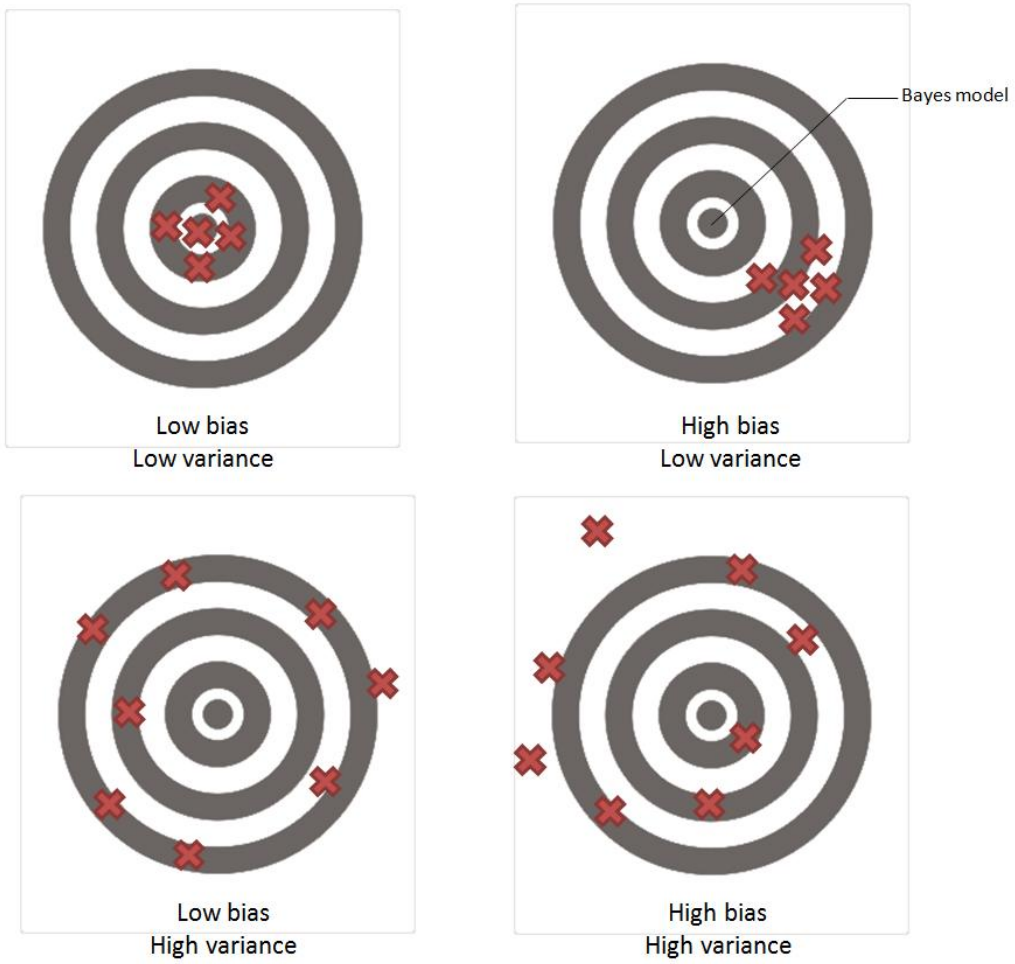
Résultats du mode ET-FL

- Relation exactitude – taille du dictionnaire

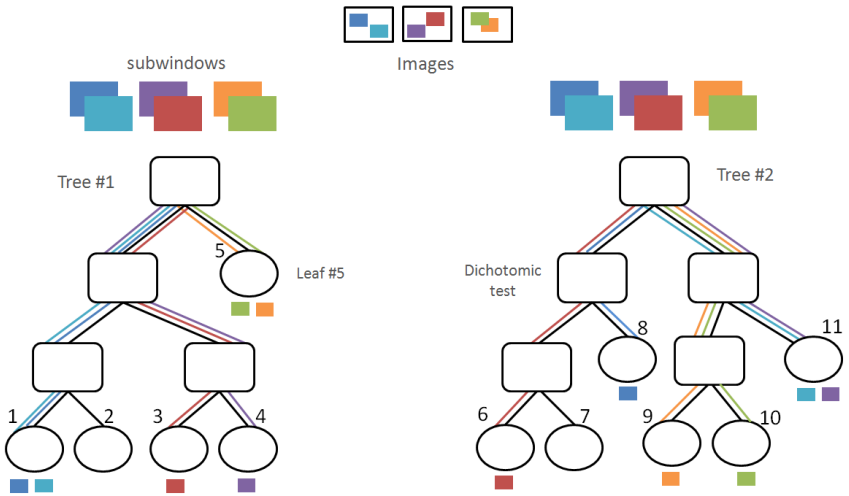


Illustrations

Bias/variance illustration

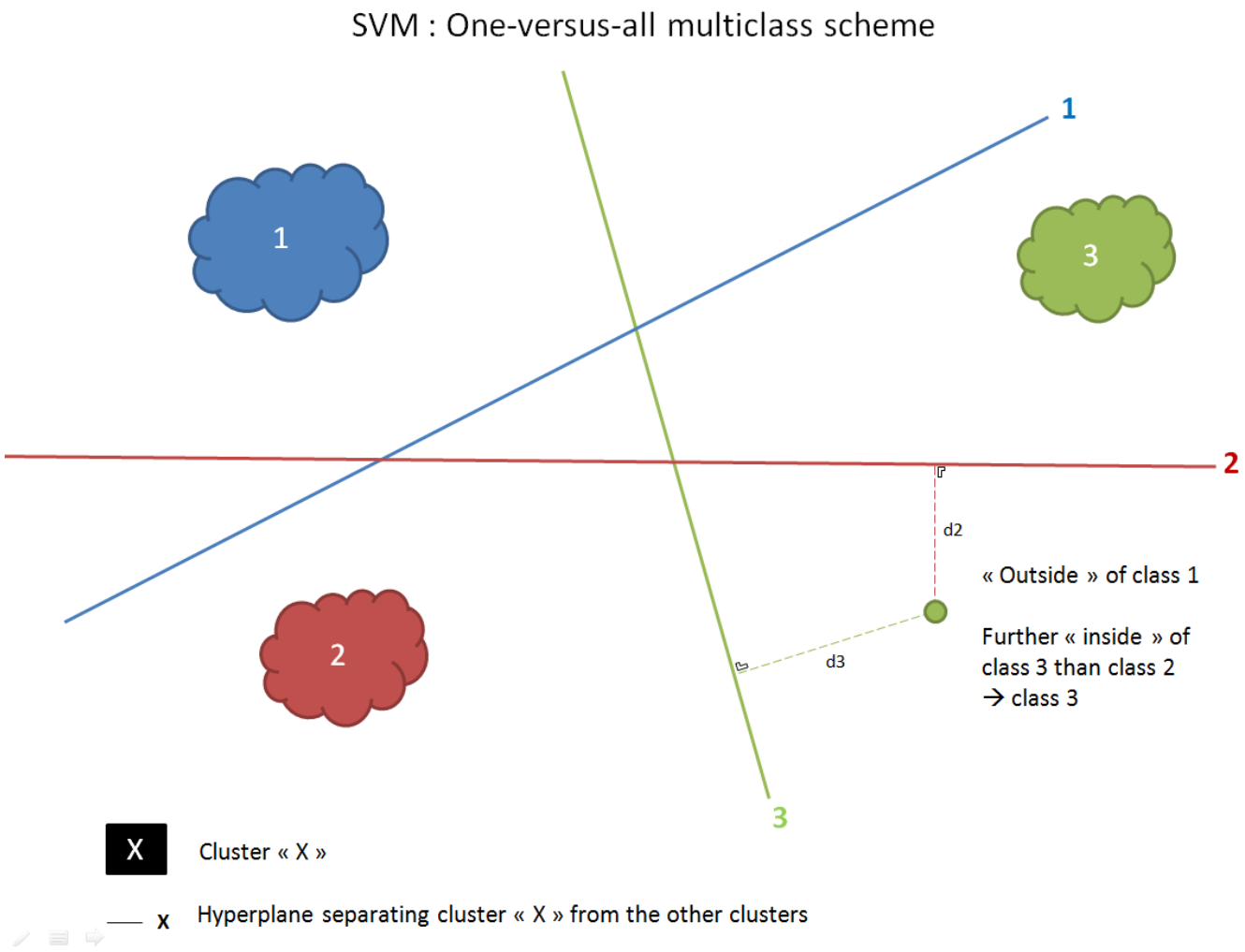


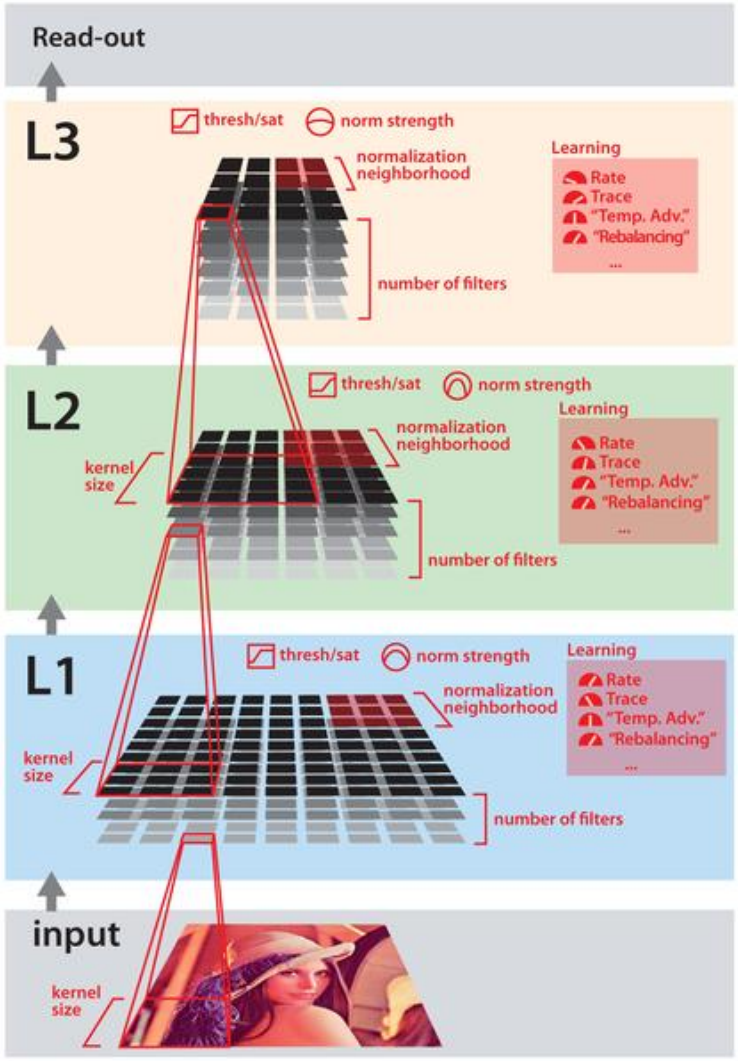
ET-FL : construction of the histogram of visual words :
Forest part



ET-FL : construction of the histogram of visual words :
Histogram part

		Leaves										
		1	2	3	4	5	6	7	8	9	10	11
Image 1	Sw. 1	I							I			
	Sw. 2	I										I
	Final word	2	0	0	0	0	0	0	1	0	0	1
Image 2	Sw. 1			I			I					
	Sw. 2				I							I
	Final word	0	0	1	1	0	1	0	0	0	0	1
Image 3	Sw. 1					I					I	
	Sw. 2					I				I		
	Final word	0	0	0	0	2	0	0	0	1	1	0


SVM : One-versus-all multiclass scheme
1
2
3
1
2
3
d2
d3
« Outside » of class 1
Further « inside » of class 3 than class 2
→ class 3
X
Cluster « X »
x
Hyperplane separating cluster « X » from the other clusters


Read-out
L3
thresh/sat
norm strength
normalization neighborhood
number of filters
Learning
Rate
Trace
"Temp. Adv."
"Rebalancing"
...
L2
thresh/sat
norm strength
normalization neighborhood
kernel size
number of filters
Learning
Rate
Trace
"Temp. Adv."
"Rebalancing"
...
L1
thresh/sat
norm strength
normalization neighborhood
kernel size
number of filters
Learning
Rate
Trace
"Temp. Adv."
"Rebalancing"
...
input
kernel size

Convolutional product

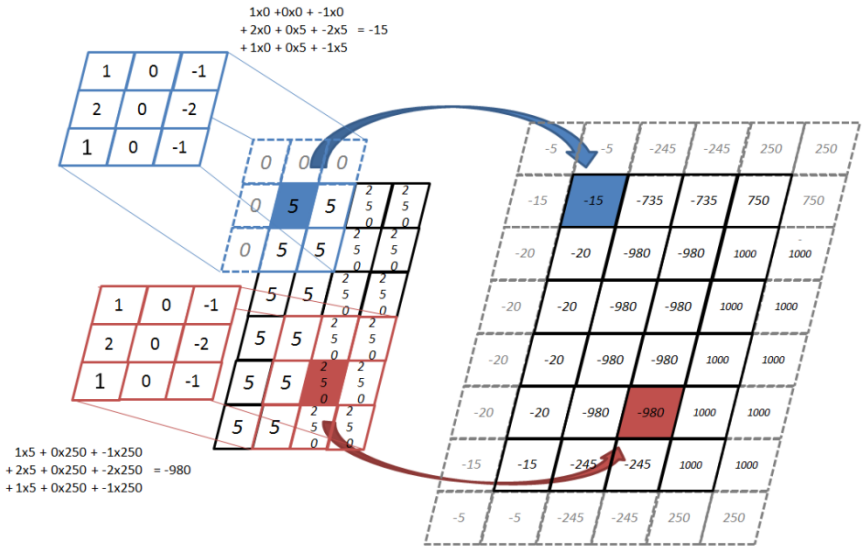
5	5	250	250
5	5	250	250
5	5	250	250
5	5	250	250
5	5	250	250
5	5	250	250

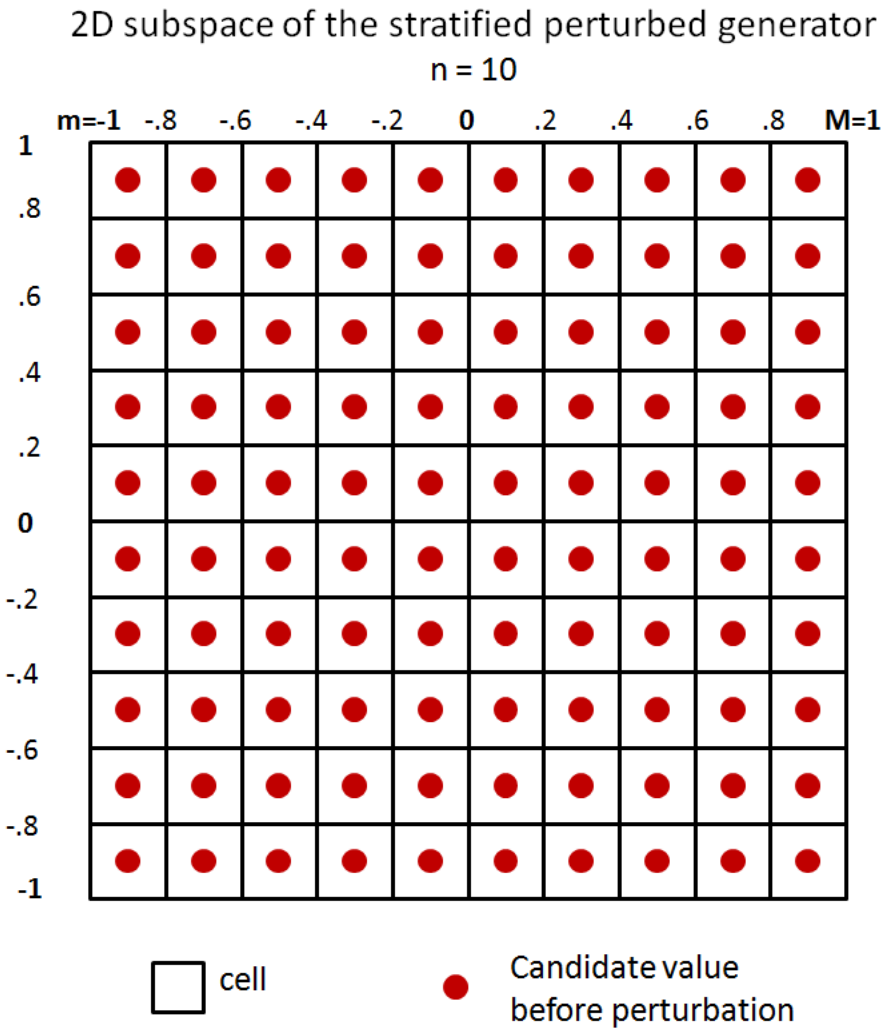
*

1	0	-1
2	0	-2
1	0	-1

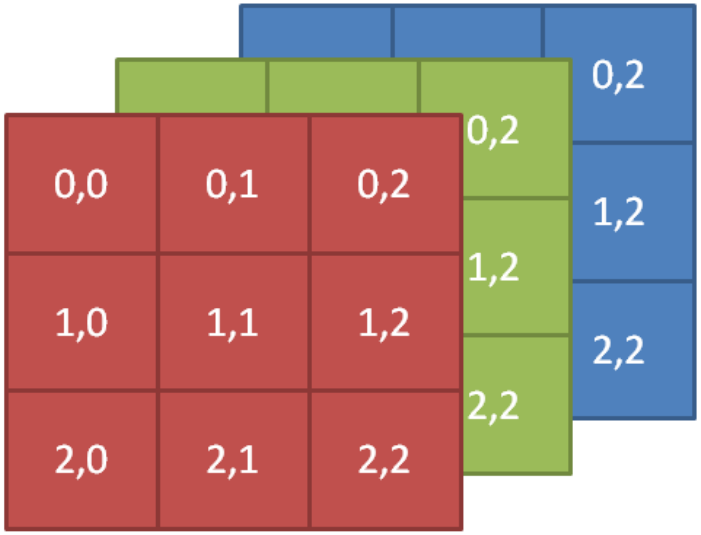
=

-15	-735	-735	750
-20	-980	-980	1000
-20	-980	-980	1000
-20	-980	-980	1000
-20	-980	-980	1000
-15	-735	-735	750

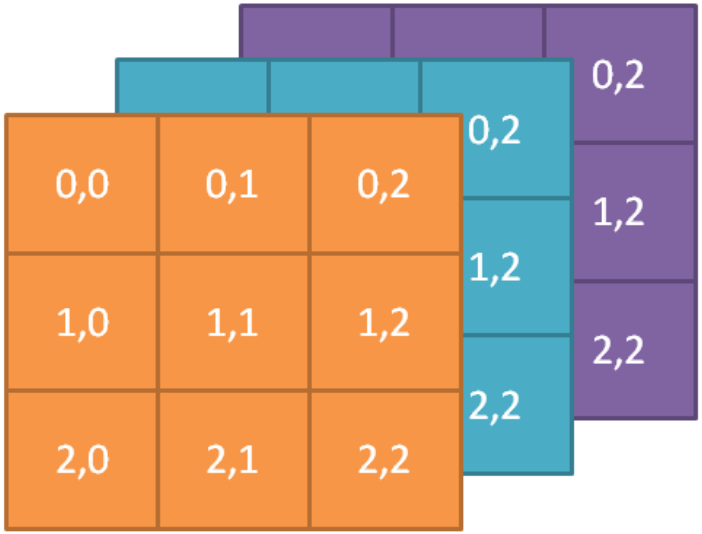



2D subspace of the stratified perturbed generator
n = 10
m=-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
M=1
1
.8
.6
.4
.2
0
-.2
-.4
-.6
-.8
-1
cell
Candidate value
before perturbation

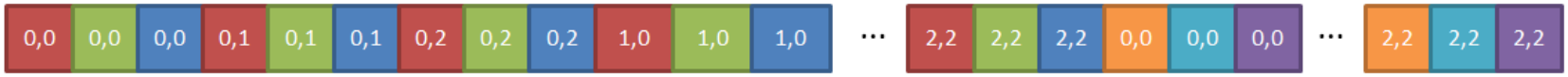
Subwindow description by raw pixels



Subwindow X of image Y
after a given filter and pooling



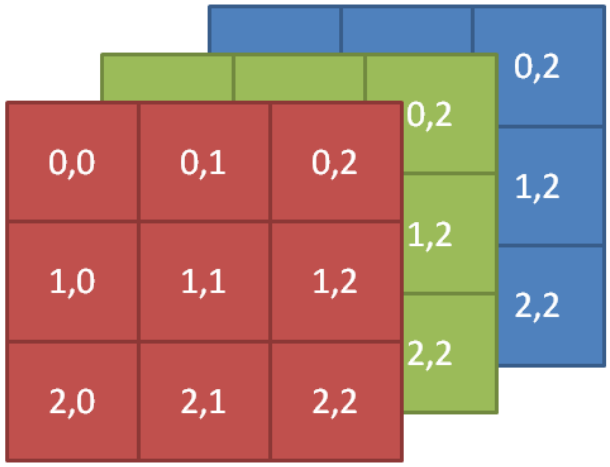
Subwindow X of image Y
after the next filter and pooling



Corresponding feature vector

Image description by raw pixels with compression

1 colors by pixel

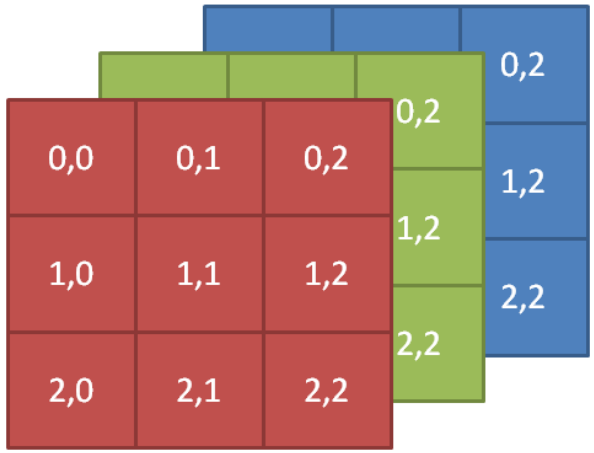


Original image

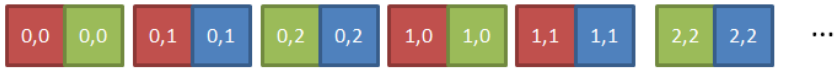
Corresponding feature vector

Image description by raw pixels with compression

2 colors by pixel



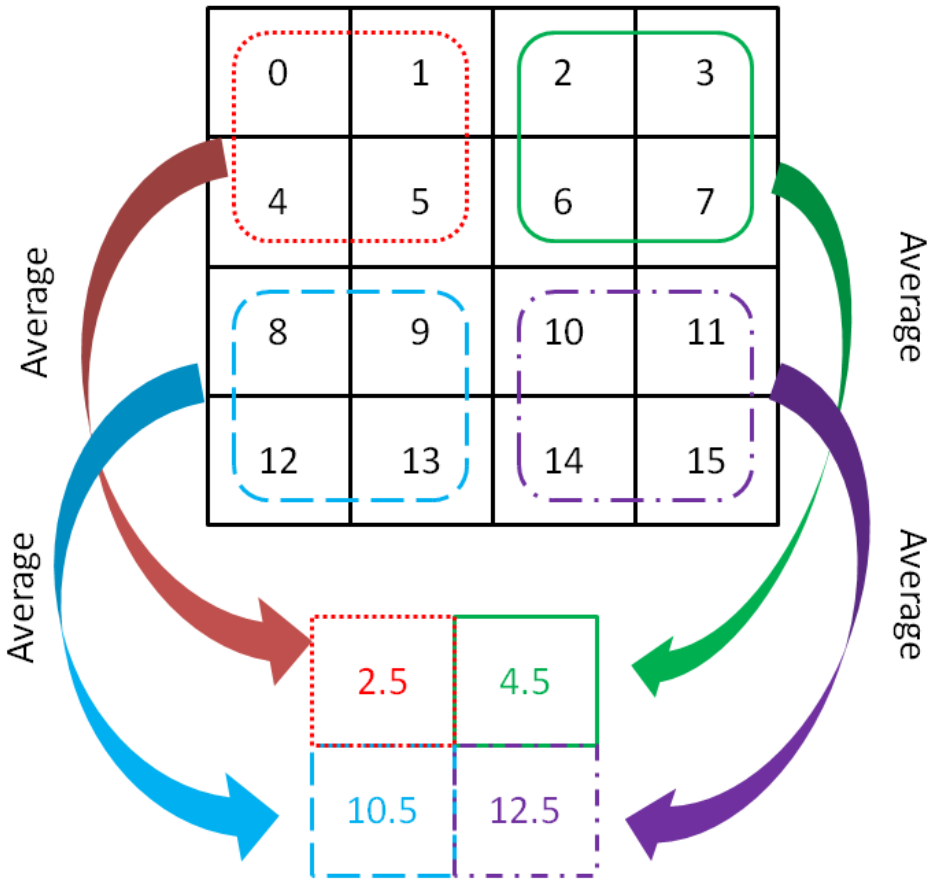
Original image



Corresponding feature vector

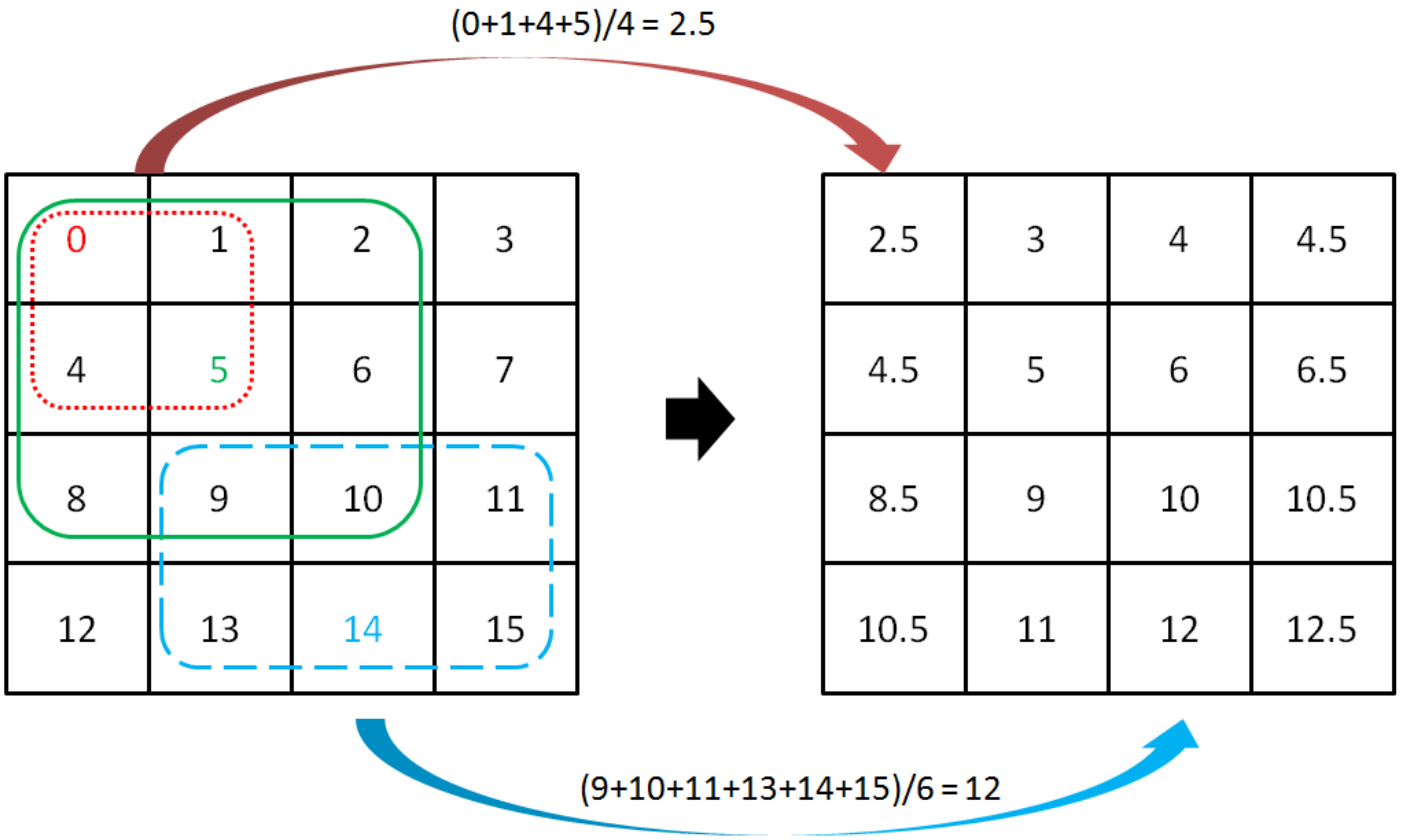
Spatial pooling by aggregation

2x2 averaging neighborhood

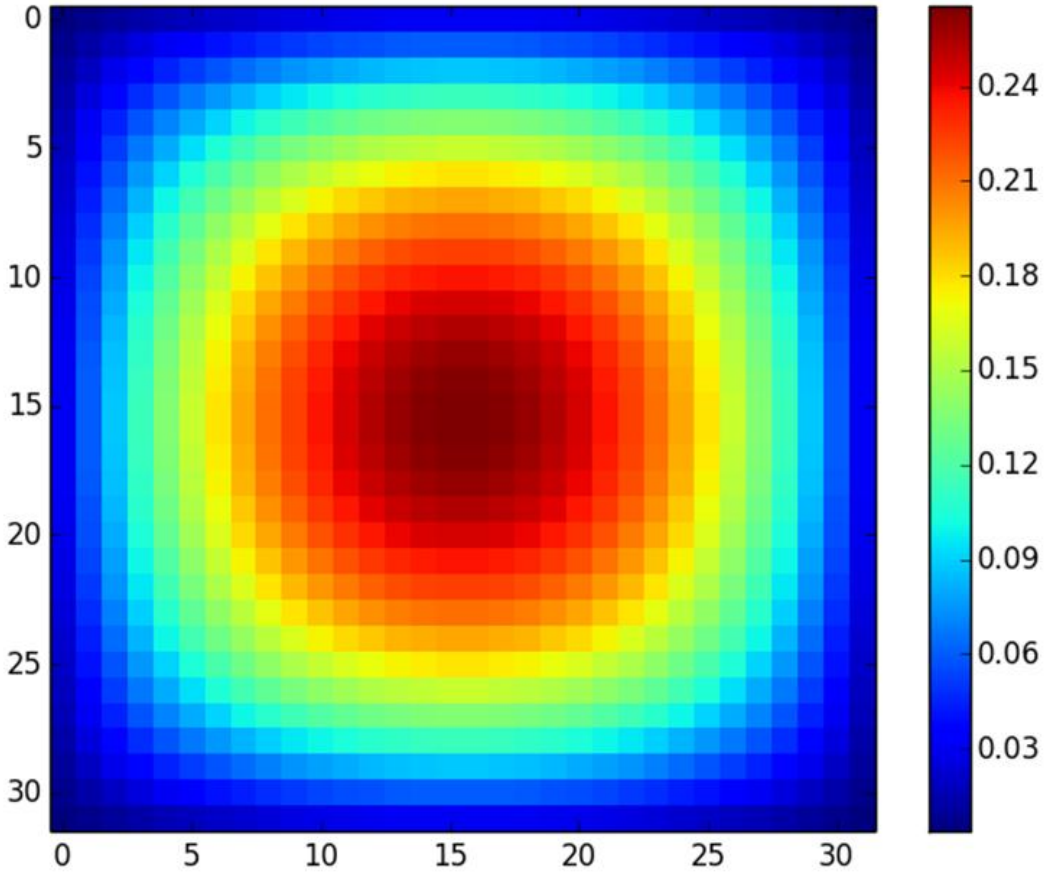


Spatial pooling by moving windows

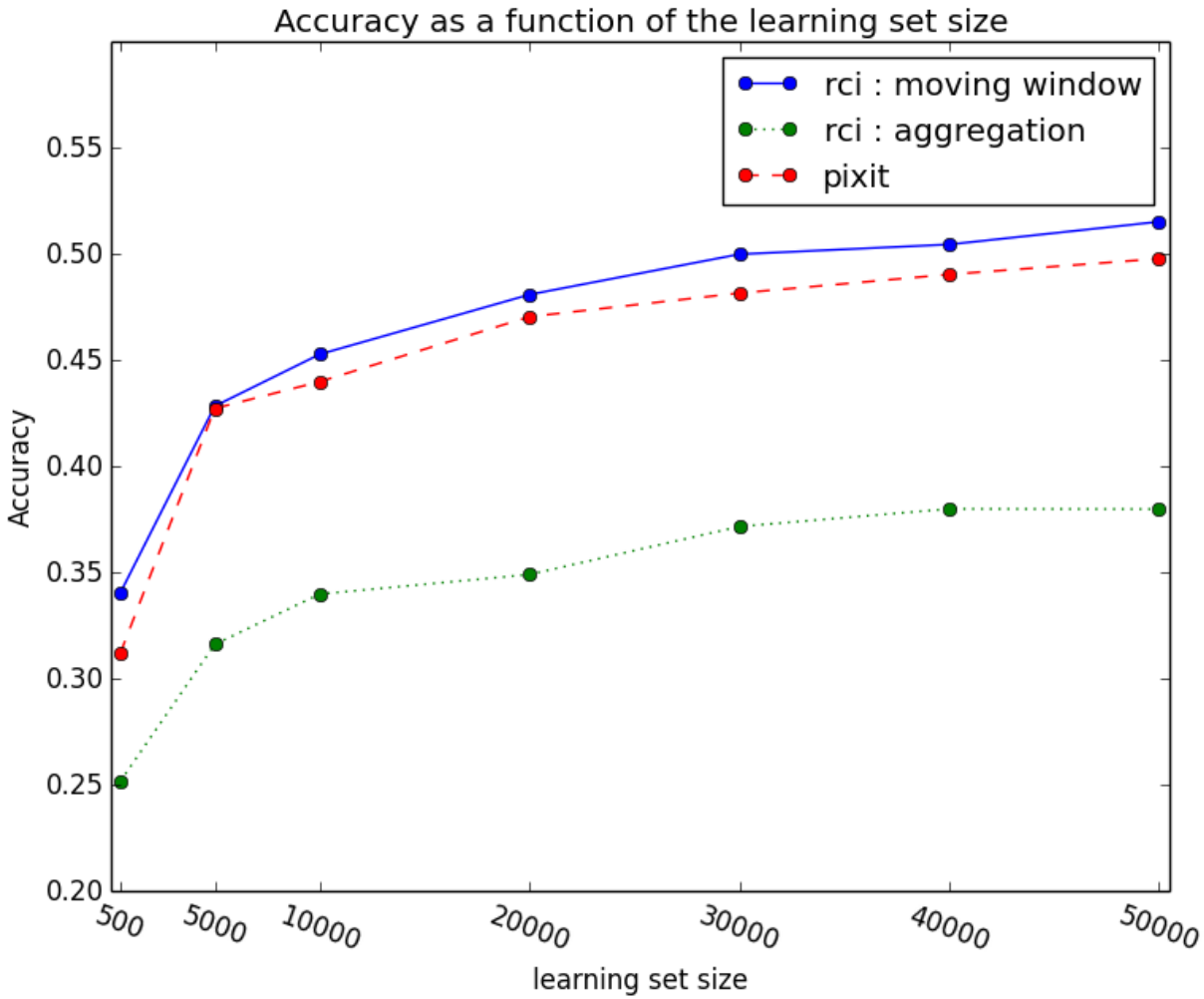
3x3 averaging windows

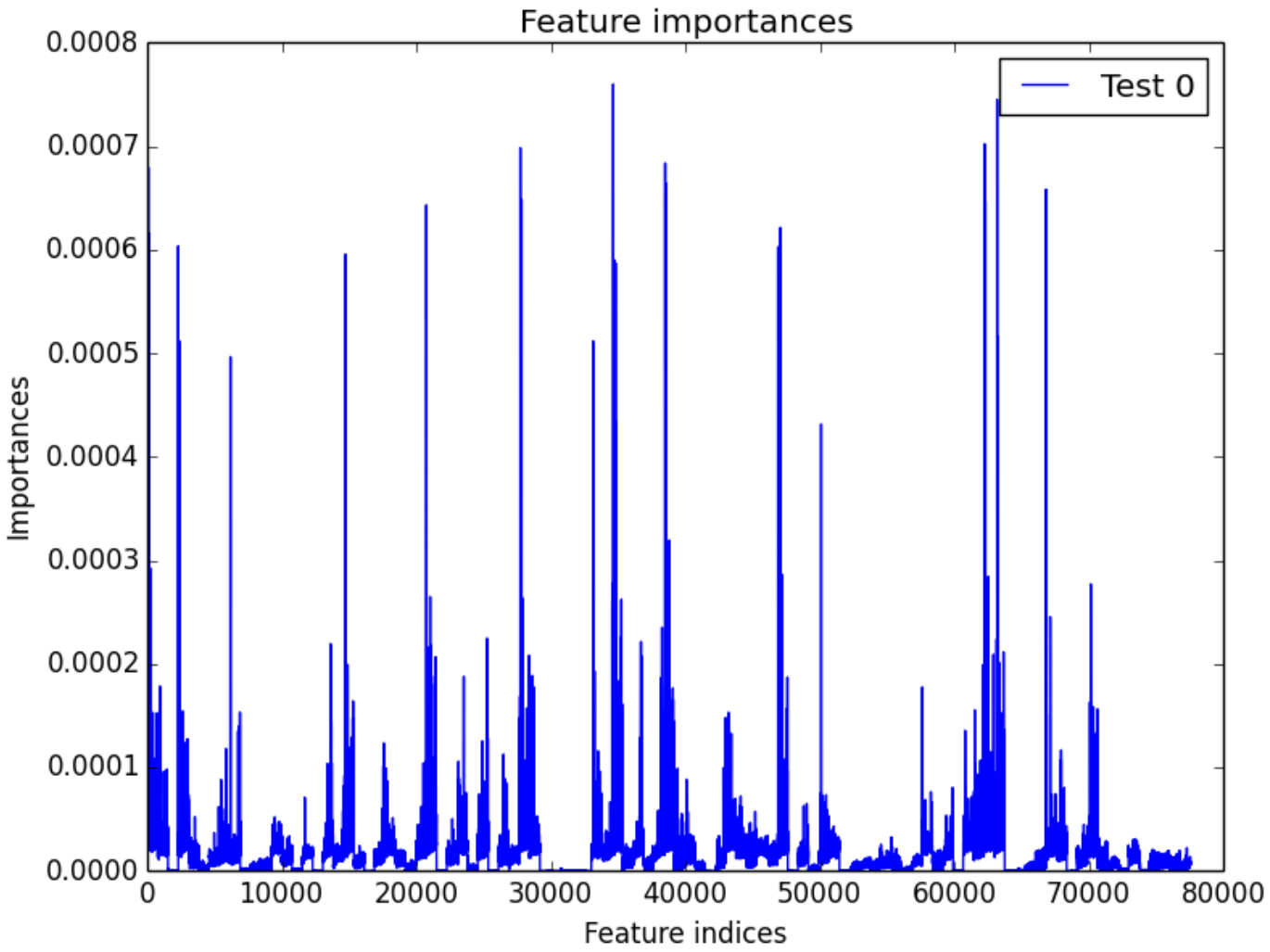


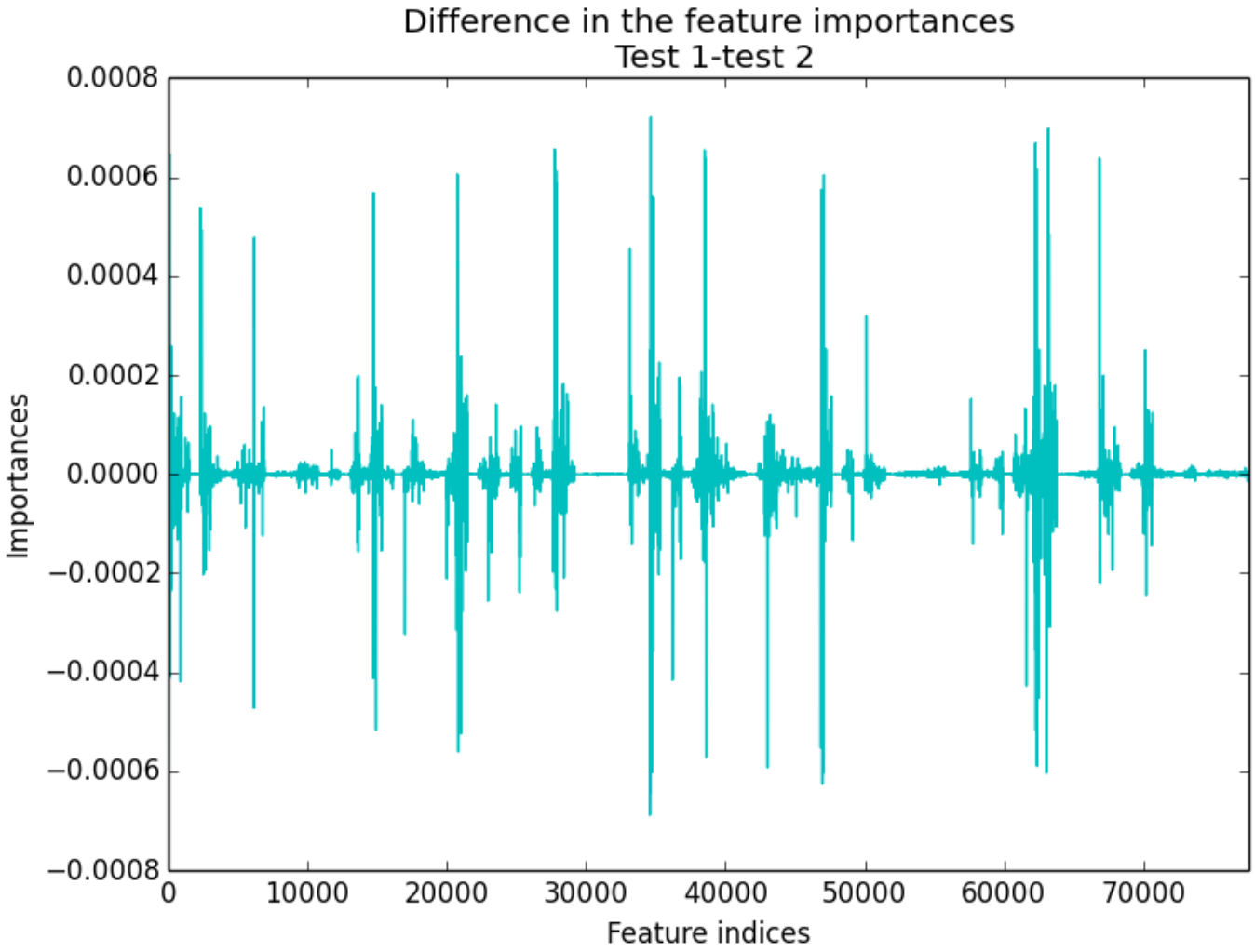
Probability per pixel of being included in a subwindow

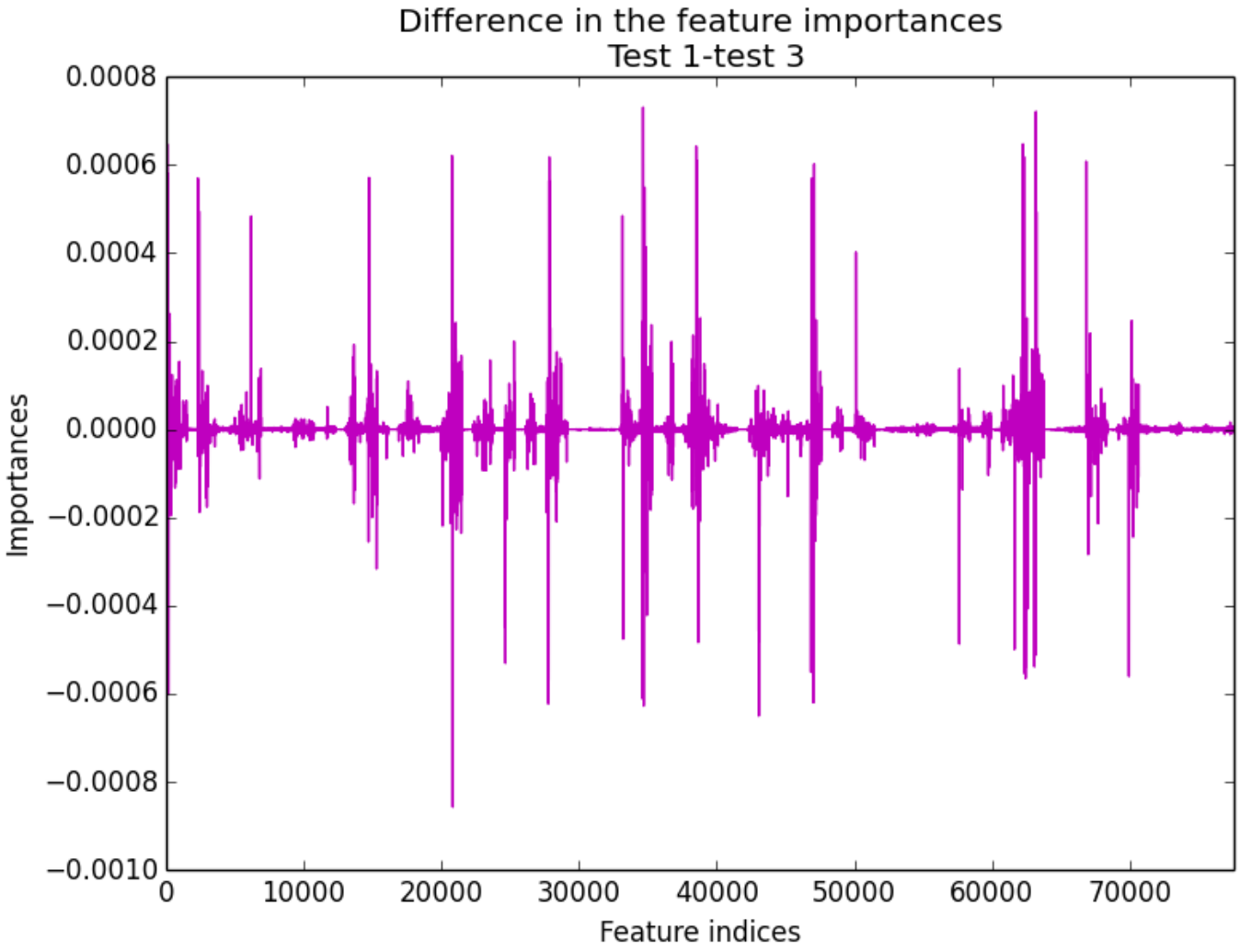


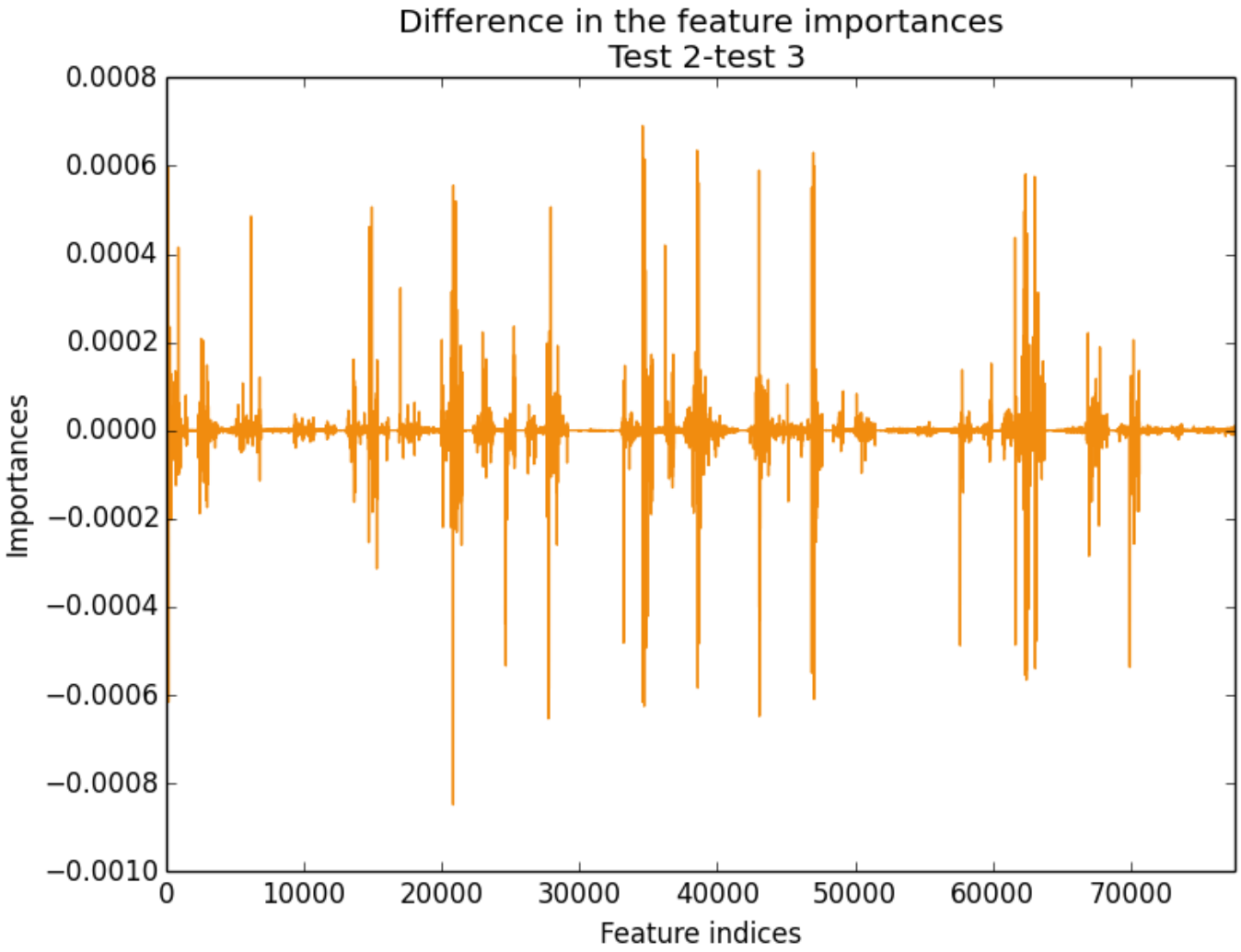
Graphes

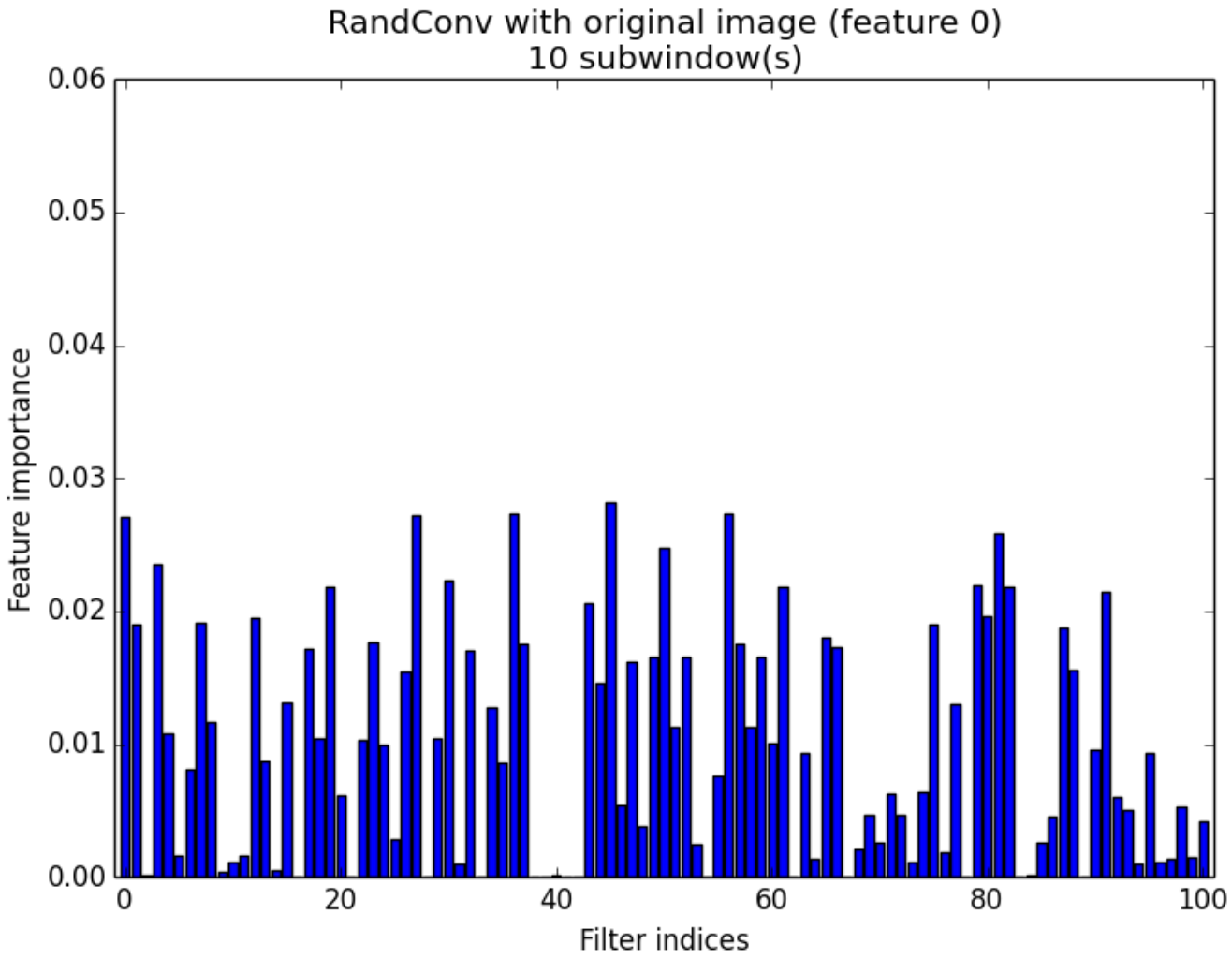

Accuracy as a function of the learning set size
rci : moving window
rci : aggregation
pixit
Accuracy
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
500
5000
10000
20000
30000
40000
50000
learning set size

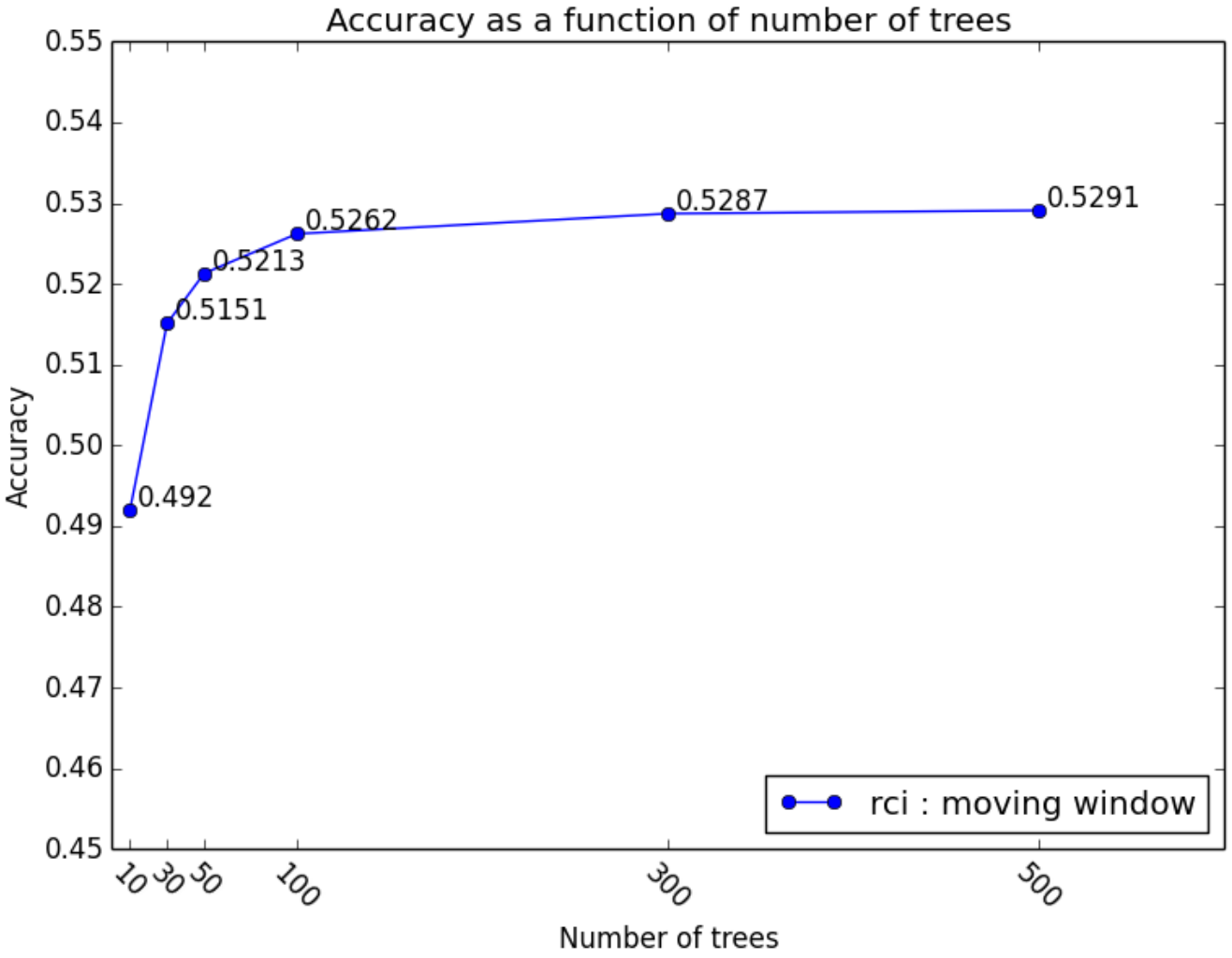

Feature importances
Test 0
Importances
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0
10000
20000
30000
40000
50000
60000
70000
80000
Feature indices

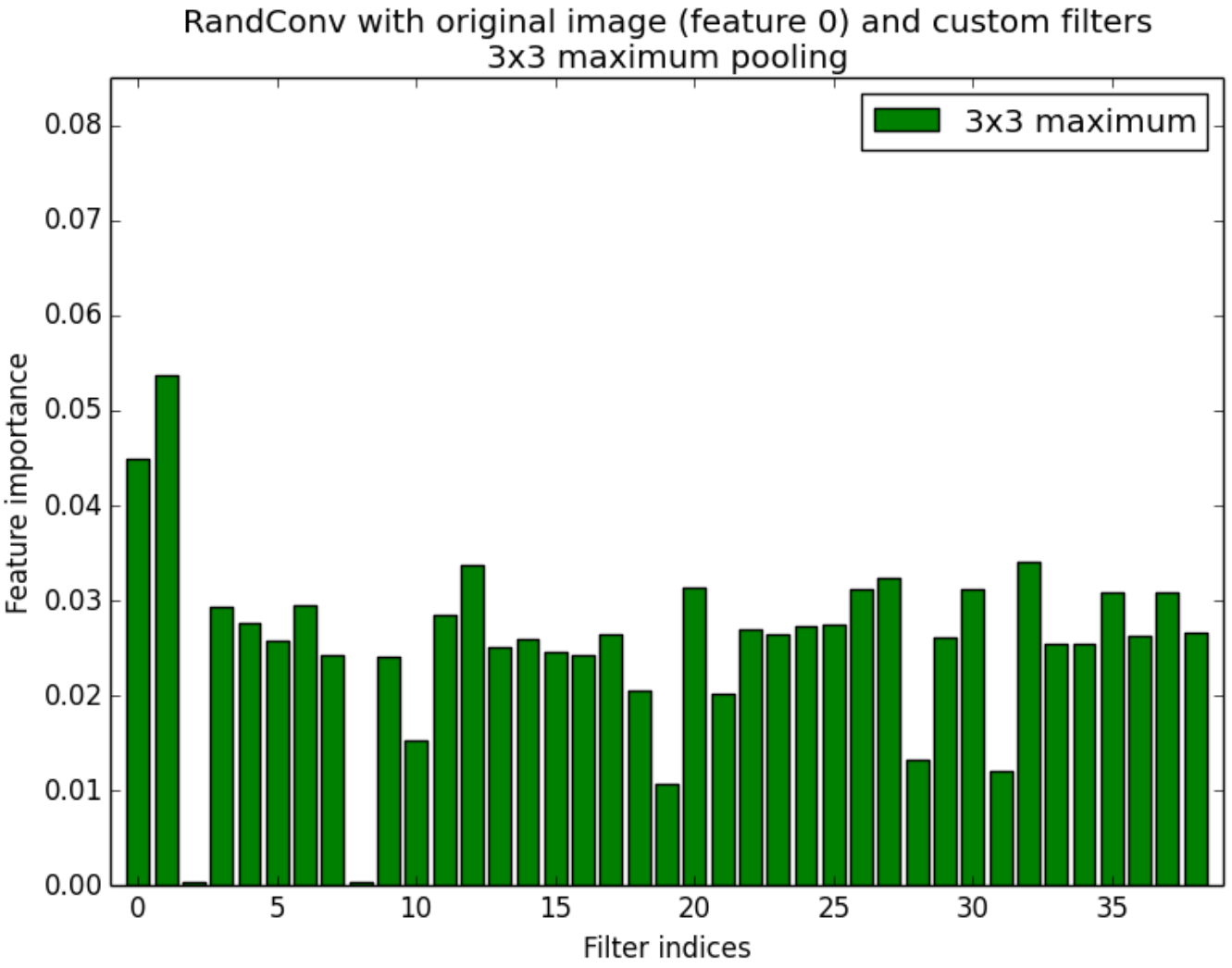

Difference in the feature importances
Test 1-test 2
Importances
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
−0.0008
0
10000
20000
30000
40000
50000
60000
70000
Feature indices

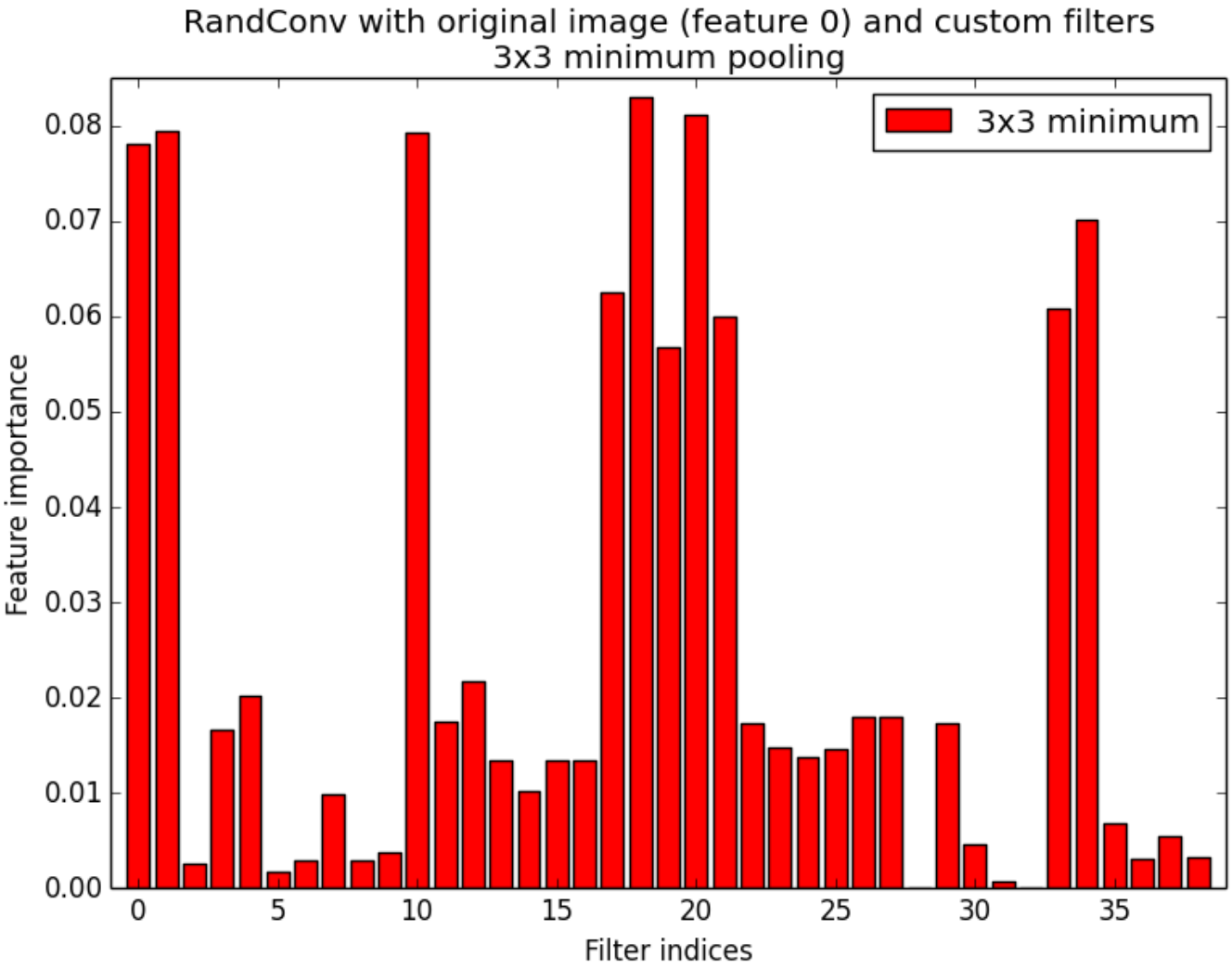

Difference in the feature importances
Test 1-test 3
Importances
Feature indices
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
−0.0008
−0.0010
0
10000
20000
30000
40000
50000
60000
70000

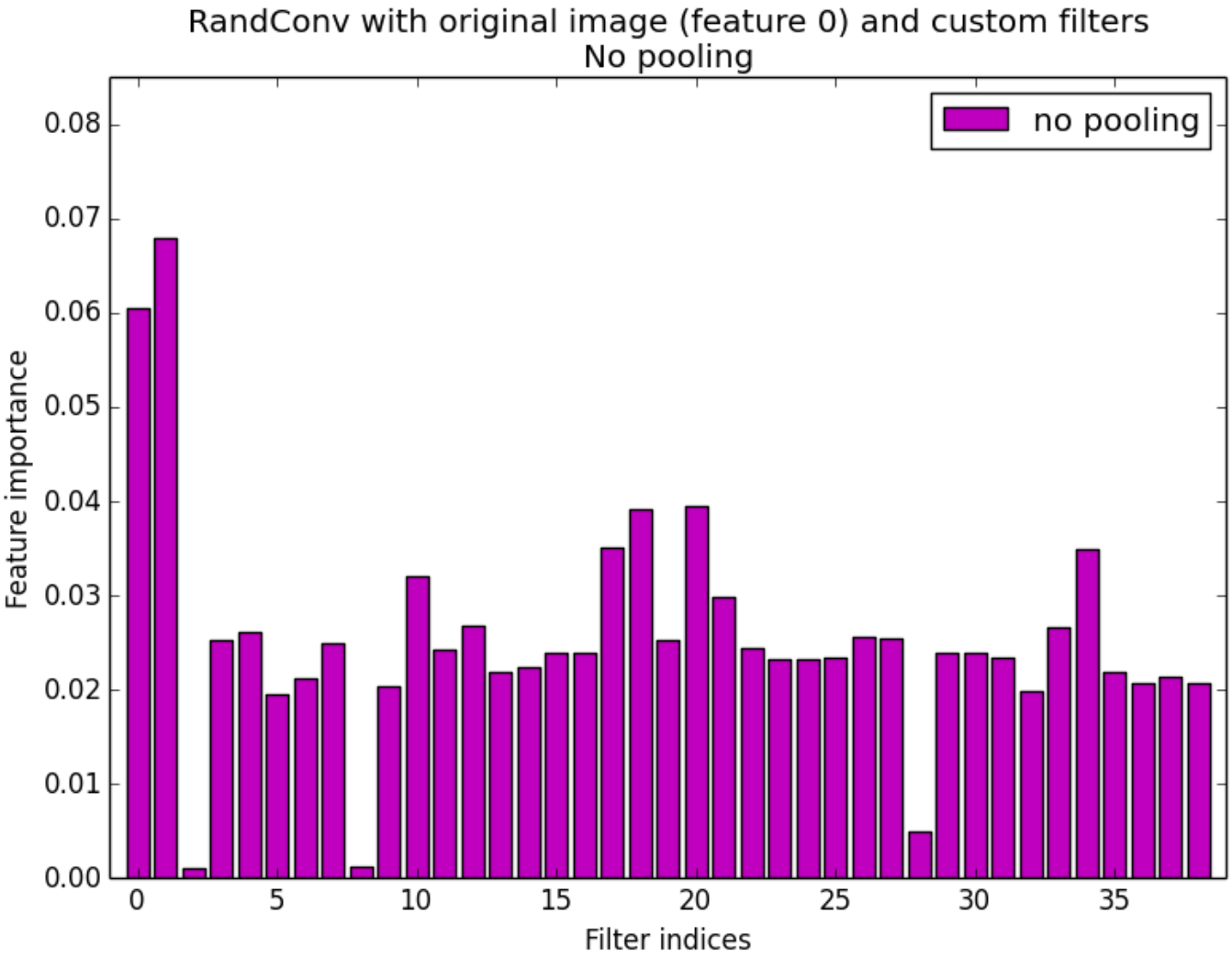

Difference in the feature importances
Test 2-test 3
Importances
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
−0.0008
−0.0010
0
10000
20000
30000
40000
50000
60000
70000
Feature indices

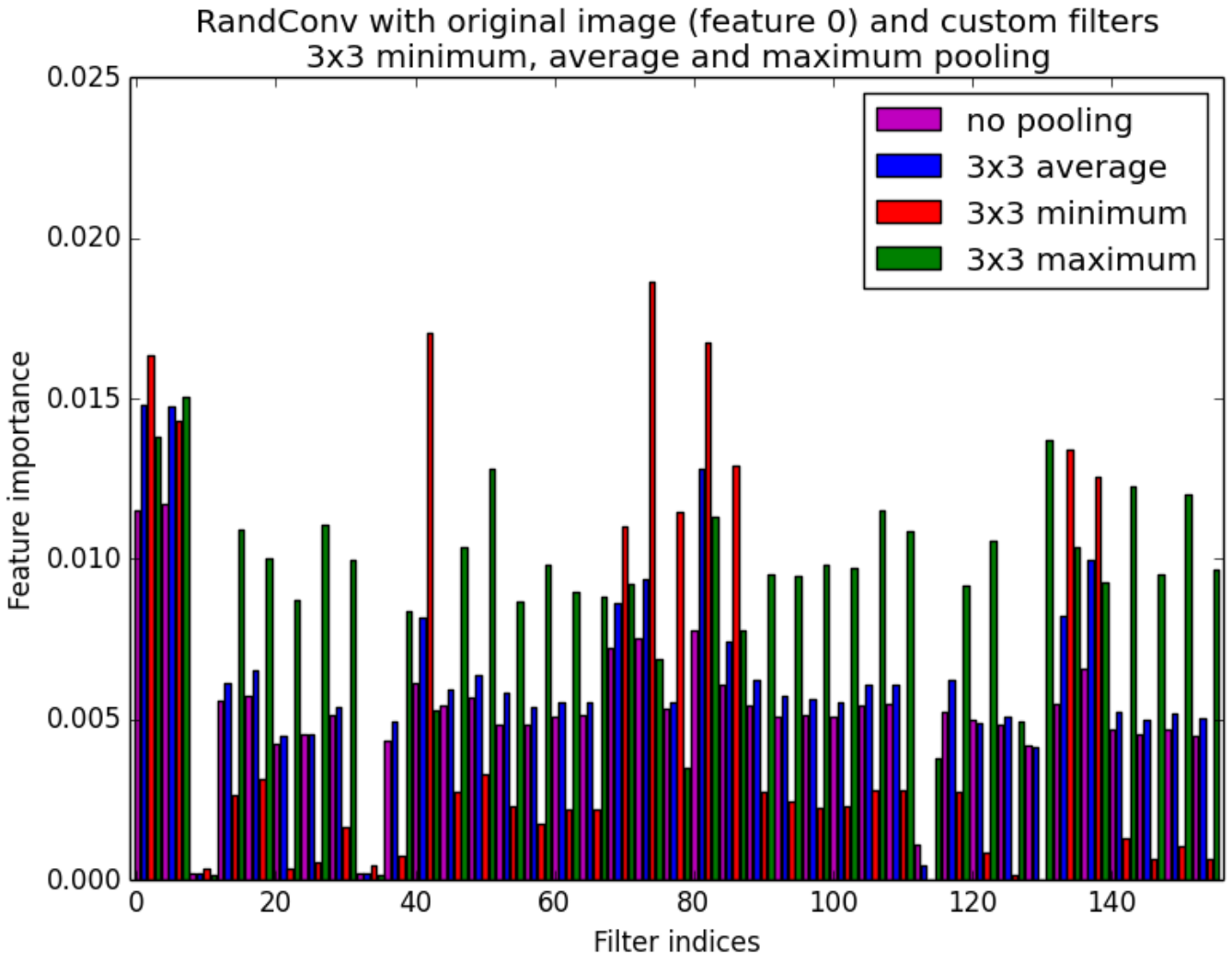

RandConv with original image (feature 0)
10 subwindow(s)
Feature importance
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
Filter indices

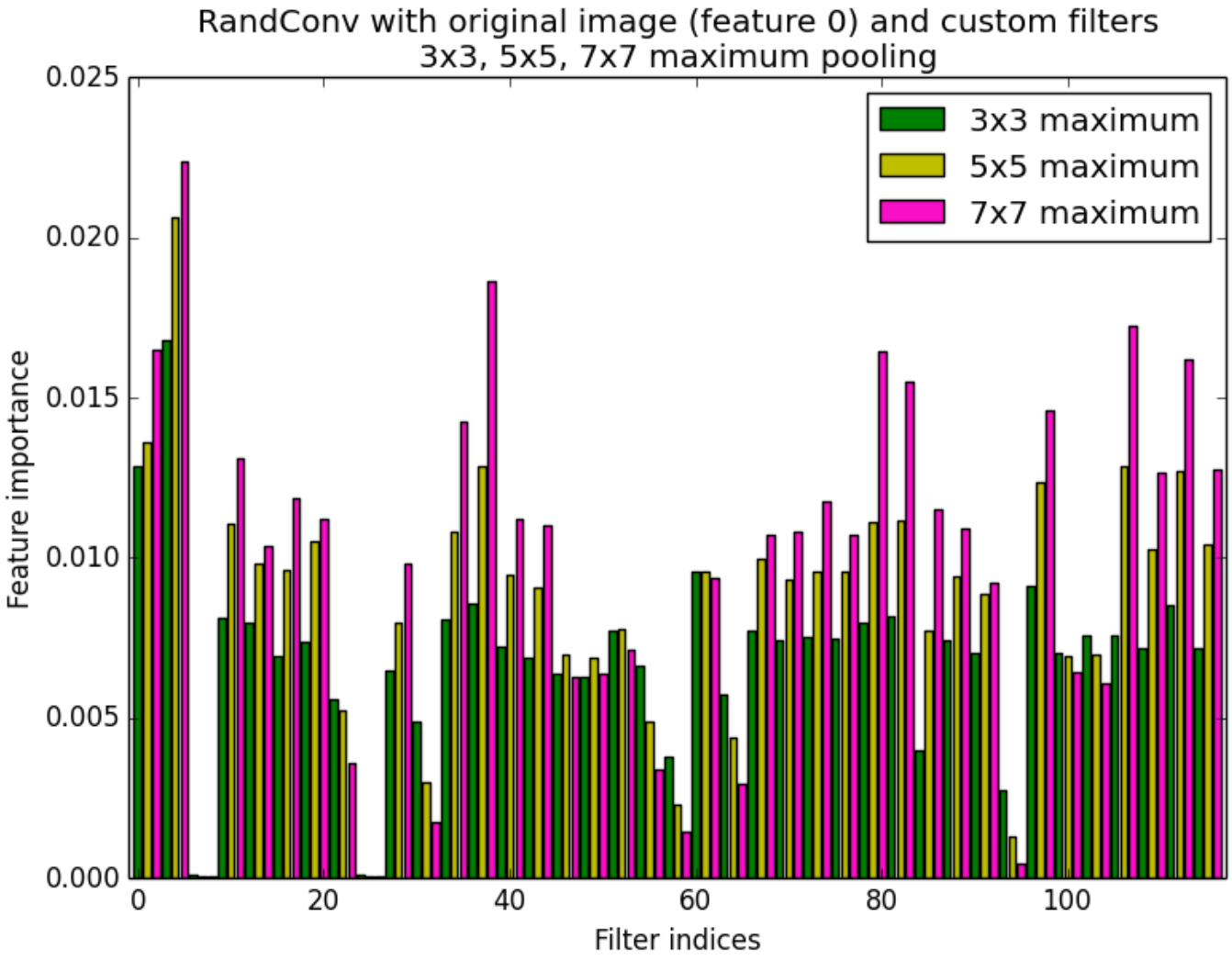

Accuracy as a function of number of trees
0.55
0.54
0.53
0.52
0.51
0.50
0.49
0.48
0.47
0.46
0.45
Accuracy
0.492
0.5151
0.5213
0.5262
0.5287
0.5291
10
30
50
100
300
500
Number of trees
rci : moving window

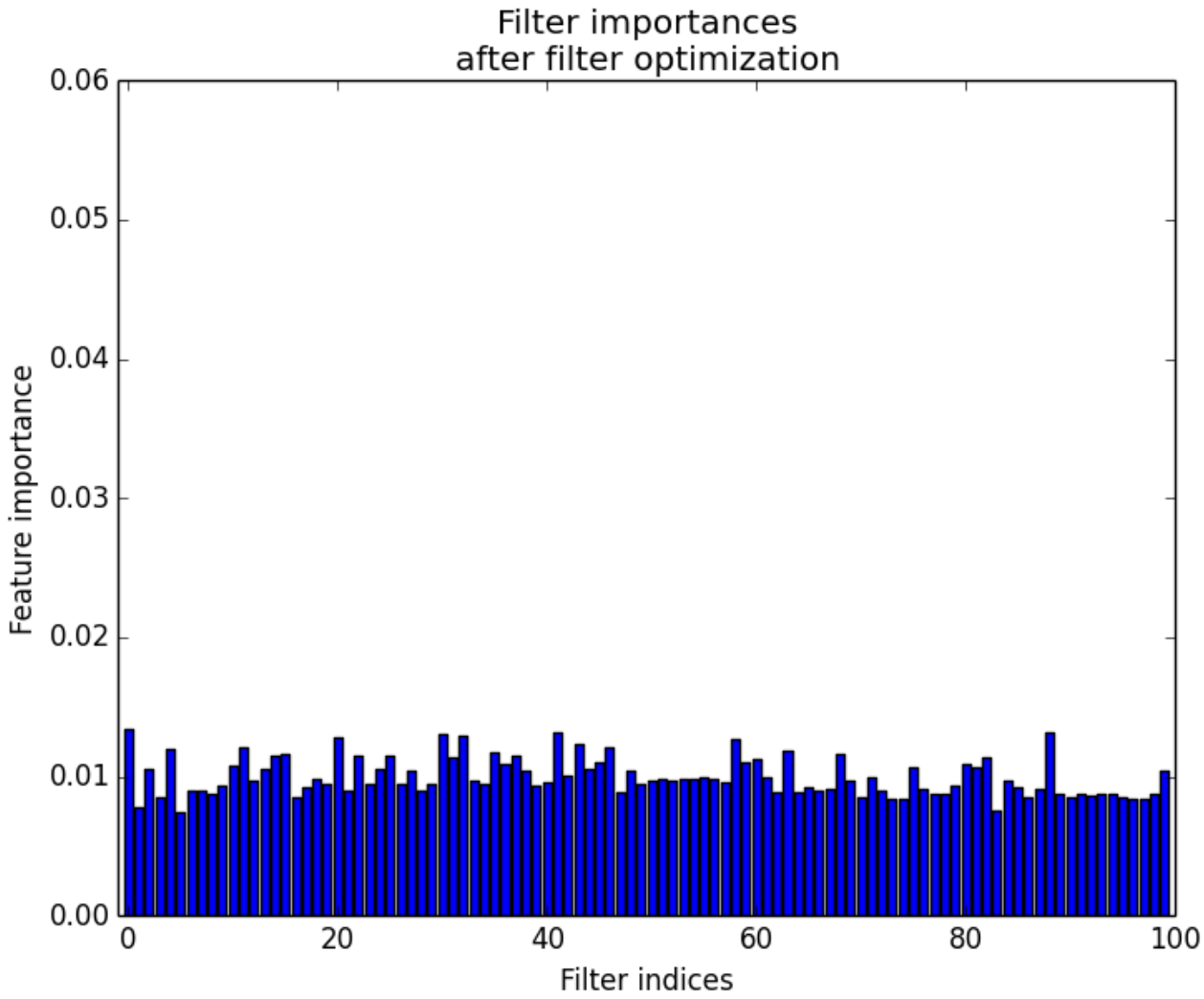

RandConv with original image (feature 0) and custom filters
3x3 maximum pooling
3x3 maximum
Feature importance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0
5
10
15
20
25
30
35
Filter indices

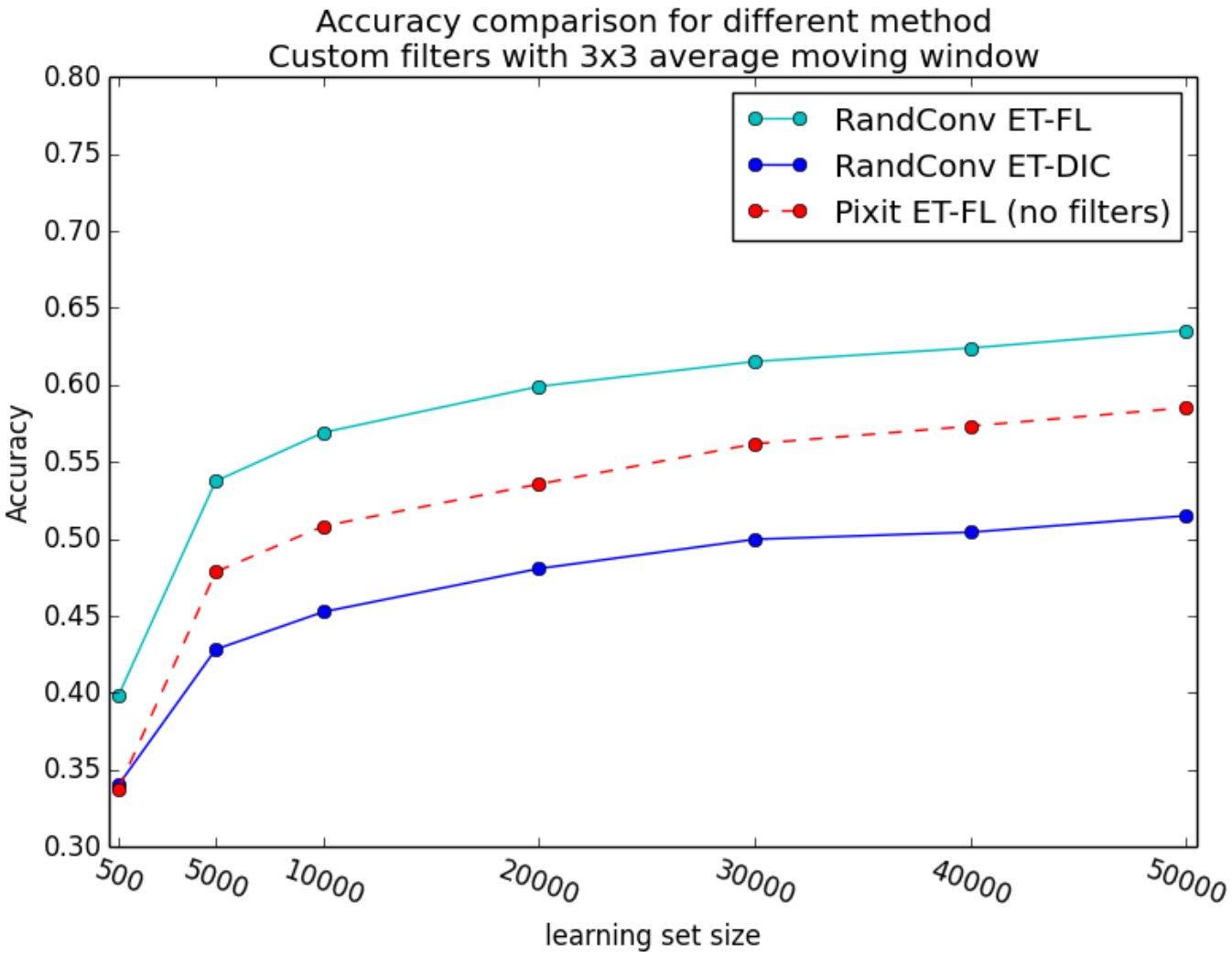

RandConv with original image (feature 0) and custom filters
3x3 minimum pooling
3x3 minimum
Feature importance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0
5
10
15
20
25
30
35
Filter indices

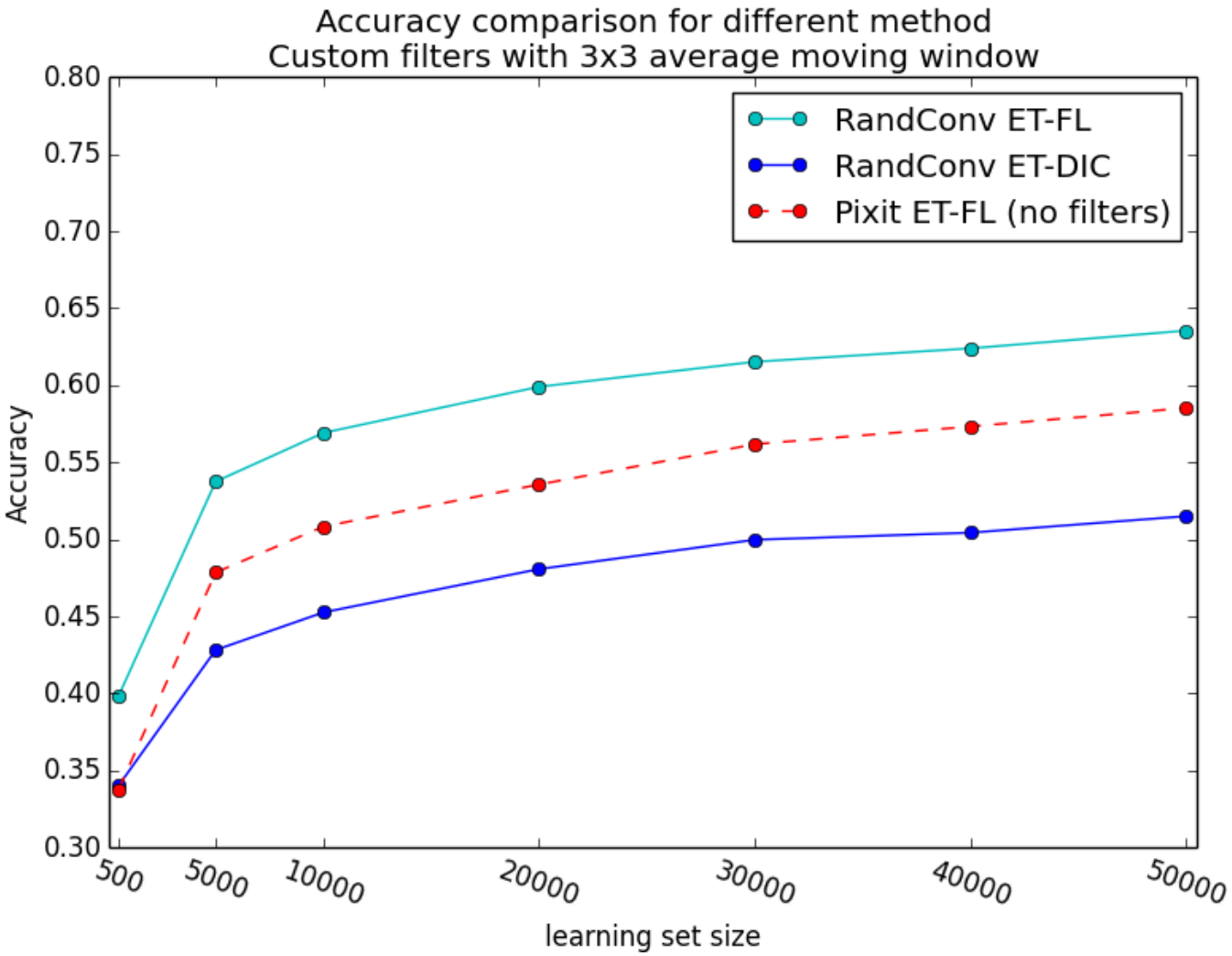

RandConv with original image (feature 0) and custom filters
No pooling
no pooling
Feature importance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0
5
10
15
20
25
30
35
Filter indices

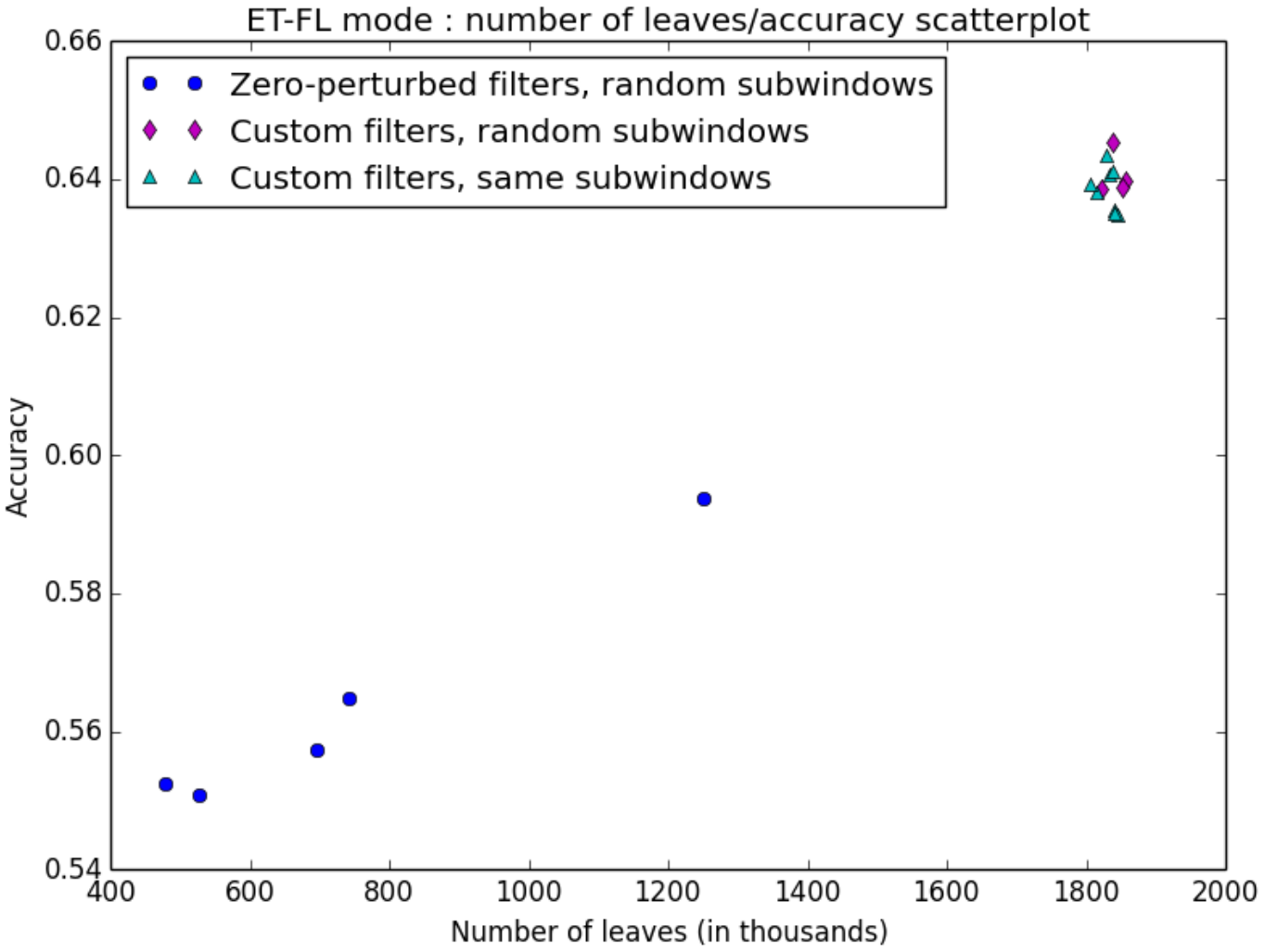

RandConv with original image (feature 0) and custom filters
3x3 minimum, average and maximum pooling
no pooling
3x3 average
3x3 minimum
3x3 maximum
Feature importance
Filter indices
0.025
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
120
140

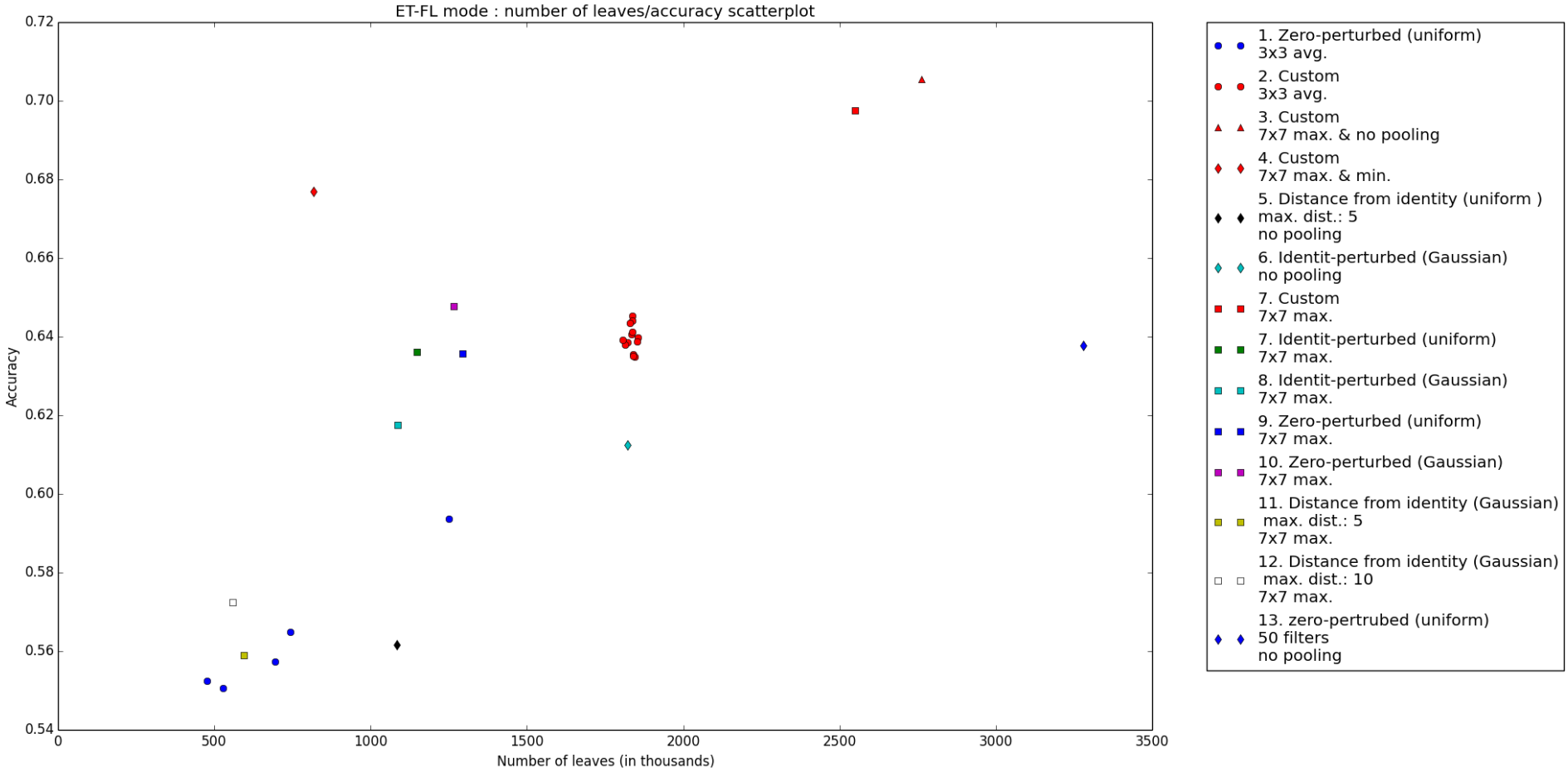

RandConv with original image (feature 0) and custom filters
3x3, 5x5, 7x7 maximum pooling
3x3 maximum
5x5 maximum
7x7 maximum
Feature importance
0.025
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
Filter indices

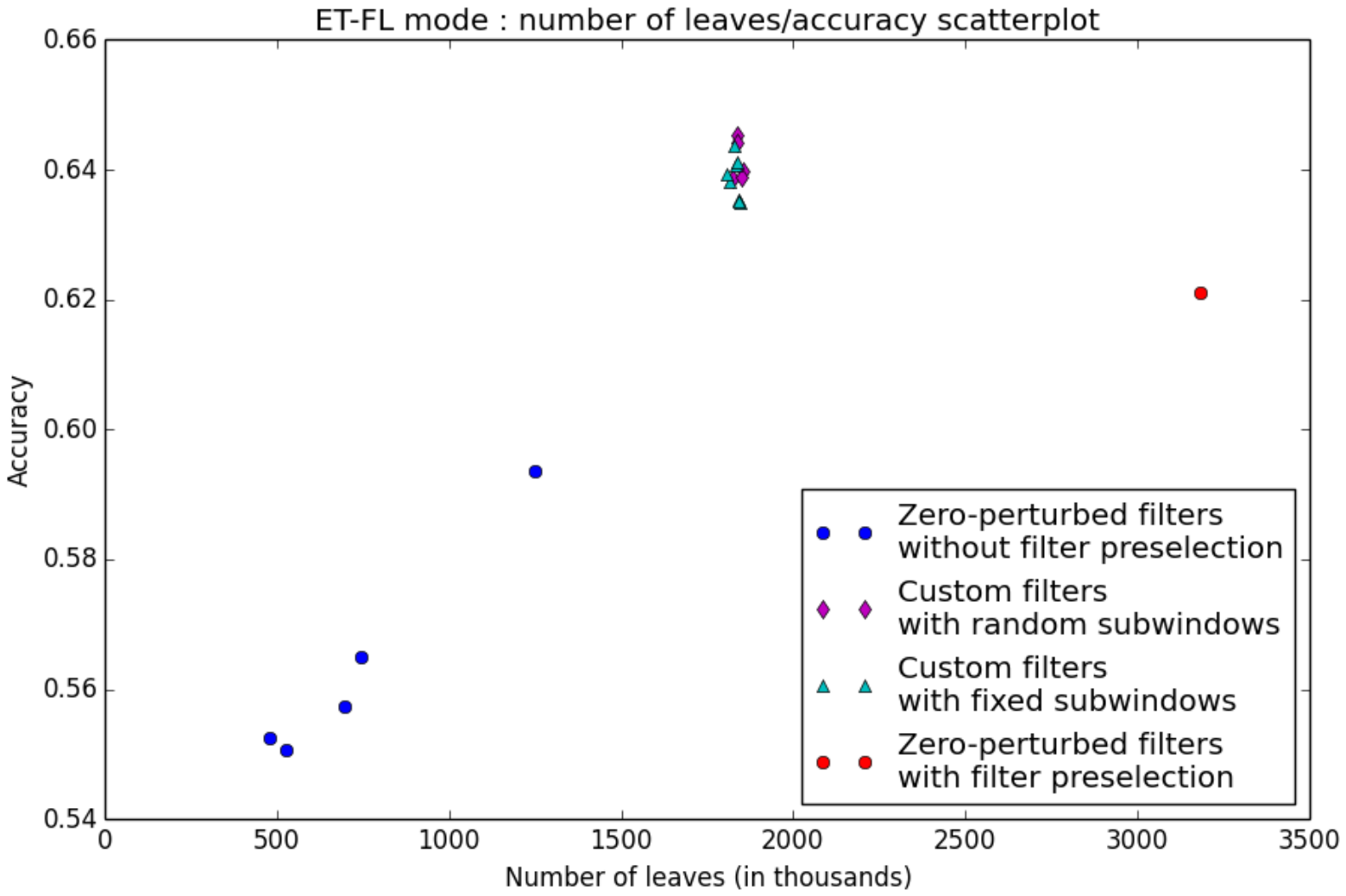

Filter importances
after filter optimization
Feature importance
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
Filter indices


Accuracy comparison for different method
Custom filters with 3x3 average moving window
RandConv ET-FL
RandConv ET-DIC
Pixit ET-FL (no filters)
Accuracy
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
500
5000
10000
20000
30000
40000
50000
learning set size


Accuracy comparison for different method
Custom filters with 3x3 average moving window
RandConv ET-FL
RandConv ET-DIC
Pixit ET-FL (no filters)
Accuracy
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
500
5000
10000
20000
30000
40000
50000
learning set size


ET-FL mode : number of leaves/accuracy scatterplot
Zero-perturbed filters, random subwindows
Custom filters, random subwindows
Custom filters, same subwindows
Accuracy
0.66
0.64
0.62
0.60
0.58
0.56
0.54
400
600
800
1000
1200
1400
1600
1800
2000
Number of leaves (in thousands)


ET-FL mode : number of leaves/accuracy scatterplot
Accuracy
Number of leaves (in thousands)
0.72
0.70
0.68
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0
500
1000
1500
2000
2500
3000
3500
1. Zero-perturbed (uniform)
3x3 avg.
2. Custom
3x3 avg.
3. Custom
7x7 max. & no pooling
4. Custom
7x7 max. & min.
5. Distance from identity (uniform)
max. dist.: 5
no pooling
6. Identit-perturbed (Gaussian)
no pooling
7. Custom
7x7 max.
7. Identit-perturbed (uniform)
7x7 max.
8. Identit-perturbed (Gaussian)
7x7 max.
9. Zero-perturbed (uniform)
7x7 max.
10. Zero-perturbed (Gaussian)
7x7 max.
11. Distance from identity (Gaussian)
max. dist.: 5
7x7 max.
12. Distance from identity (Gaussian)
max. dist.: 10
7x7 max.
13. zero-pertrubed (uniform)
50 filters
no pooling


ET-FL mode : number of leaves/accuracy scatterplot
Accuracy
Number of leaves (in thousands)
0.66
0.64
0.62
0.60
0.58
0.56
0.54
0
500
1000
1500
2000
2500
3000
3500
Zero-perturbed filters without filter preselection
Custom filters with random subwindows
Custom filters with fixed subwindows
Zero-perturbed filters with filter preselection