

Machine Learning et Deep Learning qui
facilitent la classification des vêtements.

RAPPORT

Vêtements classification

Ziqi XU

Introduction

Ce rapport est pour objectif de classifier les vêtements selon leur type, puisque beaucoup d'utilisateur rentrent incorrectement le type du vêtement. Nous avons implémenté plusieurs méthodes de l'apprentissage automatique pour tenter de résoudre ce problème.

Machine Learning

Nous avons implémenté **SVC** et **Random Forest** pour cette partie. En utilisant **SVC** avec 'C=1' et 'linear', nous n'avons pas eu le résultat car le temps du training était trop long. Si l'on utilise 1/10 des photos (soit 6000 photos), la précision (*accuracy*) est 0.81.

Les résultats du **Random Forest** sont ci-dessous en utilisant 60000 photos dont 20% pour le test. Le temps du training était très court alors que la précision était satisfaisante pour le troisième essai.

n_estimators	depth	criterion	accuracy	Training time (min)
50	10	Gini	0.854	0.257
100	10	Gini	0.853	0.489
100	50	Gini	0.882	0.855
100	50	Entropy	0.881	1.012

Deep Learning

Nous avons constaté que les photos sont similaires à celles de MNIST, en conséquence nous avons abordé cette question avec **LeNet**, un 'Neuron Network' pour classifier les nombres.

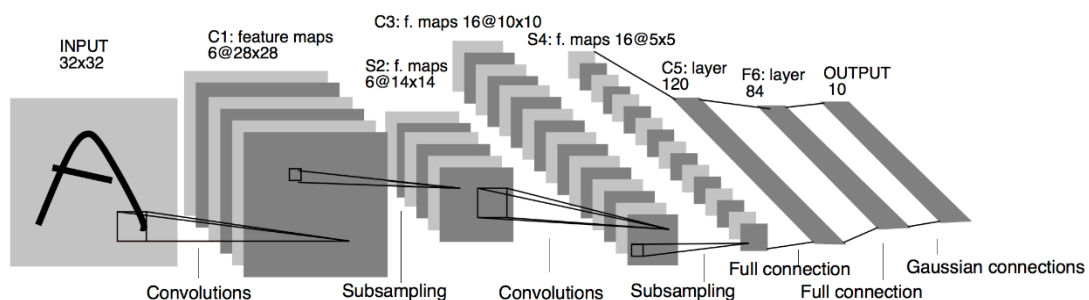


Figure 1 LeNet

Nous avons utilisé **PyTorch** pour le training (*Learning rate* 0.003). La précision était 0.91 pour **LeNet** par défaut, qui est déjà très satisfaisante par rapport à celle de Machine Learning. Nous avons également essayé le '*optimizer Adam*', qui est plus avancé (rapide, capable d'éviter '*local minima*'). La précision était 0.93 (Voir le tableau ci-dessous), alors que le temps du training était seulement une minute plus que 'SGD'.

	Epoch	Accuracy	Training time (min)
LeNet SGD	15	0.91	8.08
LeNet Adam	15	0.93	9.19
LeNet Adam (CUDA)	15	0.93	3.55

En conclusion, '**LeNet Adam**' est capable de classifier nos données du vêtement avec un temps du training raisonnable. Il est recommandé d'utiliser CUDA car il réduit significativement le temps du training.

Jeu de Données Extérieur

Les 60000 photos (Fashion MNIST) que nous avons déjà entraînées sont bien procédées, car elles sont bien visibles sans bruit du fond. En revanche, les photos soumises par les utilisateurs sont plus compliquées, qui sont moins '**idéales**' que Fashion MNIST. En conséquence, nous avons trouvé deux jeux de données pour mieux comprendre la difficulté de classification. Les photos sont en couleur et beaucoup plus grandes (majoritairement 320 * 320). Compte tenu que notre ordinateur n'est pas assez puissant, nous avons fait le training sur le premier jeu de données.

Dataset Name	Size	Class	Paper/Thèse
Apparel classification with Style	>80,000	15	ACCV 12 ¹
DeepFashion	>280,000	50	CVPR 2016 ²

¹ https://data.vision.ee.ethz.ch/cvl/lbssard/accv12/accv12_apparel-classification-with-style.pdf

² <http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html>

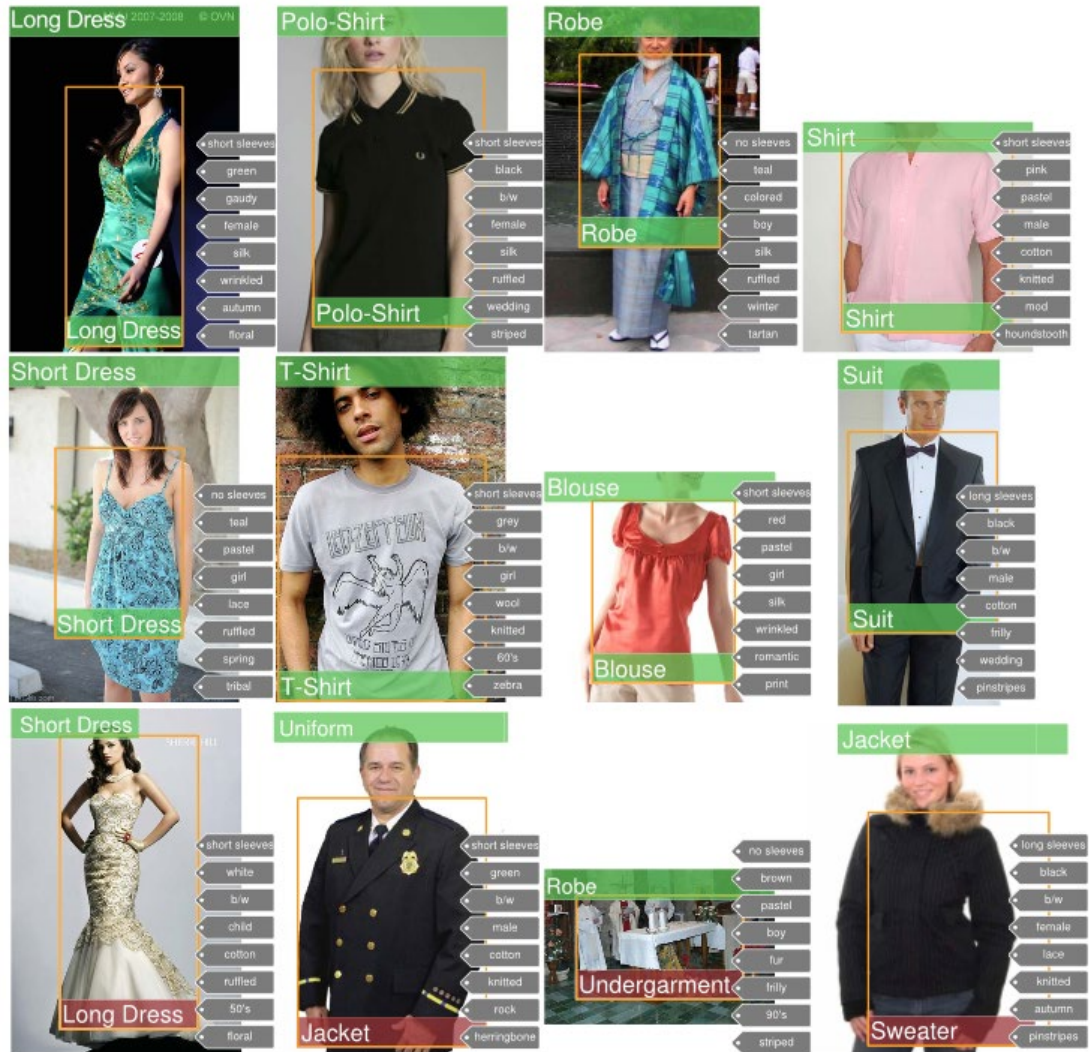


Figure 2 Exemples du jeu de données

Nous avons d'abord essayé LeNet (avec *Adam*, taille 64×64). Dans les trois premiers essais, nous avons comparé les différents paramètres du CUDA, en combinaison avec les guides sur l'Internet, il vaut mieux fixer *num_workers* (calcul parallèle) à 8 ou 16 et augmenter la taille du *batch*. La réduction du temps était très significative. En revanche, la précision (*accuracy*) était moins de 30% pour tous les cas, c'est à dire LeNet n'est pas capable de classer les photos compliquées.

	epochs	num_workers	batch	Time (min)
<i>LeNet</i>	15	-	32	42
<i>LeNet + CUDA</i>	15	-	32	19
<i>LeNet + CUDA</i>	15	4	32	5.7
<i>LeNet + CUDA</i>	50	8	128	21.6

Comparaison avec Les Thèses

Les méthodes du Machine Learning ne sont pas appliquées car le résultat de la thèse [ACCV 12](#) n'était pas satisfaisant.

Model	Accuracy
SVM	35.8%
Transfer Forest Model	41.3%

La thèse de l'équipe [Stanford](#) '*Apparel Classification using CNNs*' travaille³ sur le même jeu de données. Ils ont construit leur CNN (Figure 4) et obtenu un résultat quasiment pareil (41.1%) par rapport aux méthodes Machine Learning.

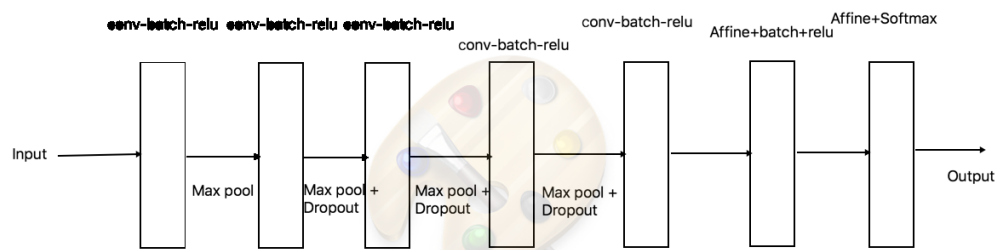


Figure 4 CNN de la thèse Stanford

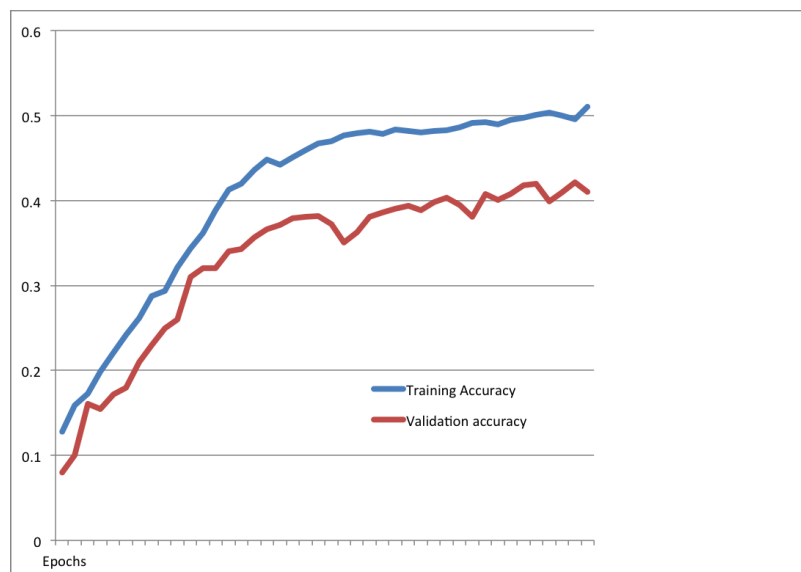


Figure 5 Résultat de la thèse Stanford

³ http://cs231n.stanford.edu/reports/2016/pdfs/286_Report.pdf

La thèse du CPVR 2016 a utilisé un autre jeu de données **DeepFashion**, les chercheurs ont conçu un CNN **FashionNet** (Figure 6) qui ressemble à VGG16. C'est un réseau très compliqué, voir la comparaison avec les autres réseaux (Figure 7), mais les précisions Top 3 et Top 5 étaient mieux que celles de ResNet18 même si les photos du **DeepFashion** sont beaucoup plus compliquées (voir la comparaison en page 4).

	Category		Notre résultat ResNet18		
	top-3	top-5	epochs	Top 3 accuracy	Top 5 accuracy
WTBI [3]	43.73	66.26	10	0.72	0.84
DARN [10]	59.48	79.58			
FashionNet+100	47.38	70.57	15	0.69	0.82
FashionNet+500	57.44	77.39			
FashionNet+Joints [34]	72.30	81.52			
FashionNet+Poselets [34]	75.34	84.87			
FashionNet (Ours)	82.58	90.17			

Figure 6 FashionNet (**DeepFashion**)

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
VGG16	528 MB	0.713	0.901	138,357,544	23
InceptionV3	92 MB	0.779	0.937	23,851,784	159
ResNet50	98 MB	0.749	0.921	25,636,712	-
Xception	88 MB	0.790	0.945	22,910,480	126
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
ResNeXt50	96 MB	0.777	0.938	25,097,128	-

The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset.

Figure 7 Comparaison

Conclusion

Les réseaux neurones (CNN) sont très efficaces pour la classification des photos. Si nous avions plus de ressources pour le training, nous aurions pu construire un CNN assez puissant que le FashionNet pour la classification des vêtements.