RATTRAPAGE · DEEP LEARNING

Reconnaître une pièce sur une photo

Classer les photos d’annonces immobilières avec un réseau de neurones convolutif et le transfer learning

Yanis Kharbouch · 5e année

Novembre 2026

LE PROBLÈME

Des photos arrivent en vrac, il faut savoir quelle pièce chacune montre

LE CONSTAT

Pour publier une annonce, les photos sont triées à la main : la cuisine, puis le salon, puis les chambres. C’est long et répétitif.

LA TÂCHE

Donner à chaque photo une étiquette parmi 7 : cuisine, salon, chambre, salle de bain, salle à manger, autre pièce, extérieur.

EN DEEP LEARNING

C’est une classification d’images supervisée : on apprend à partir d’exemples déjà étiquetés.

PLAN

Quatre étapes

1 Comment un réseau convolutif « voit » une image

2 D'ImageNet au transfer learning

3 Mon expérience sur de vraies photos de maisons

4 Limites et conclusion

Pour l’ordinateur, une image n’est qu’un tableau de nombres

Une photo de 224 × 224 pixels, en couleur (3 canaux)

150 528

valeurs en entrée

Une seule couche dense de 1 000 neurones

≈ 150 M

de poids à apprendre, et la notion de pixels voisins est perdue

Un petit filtre glisse sur l’image et cherche un motif

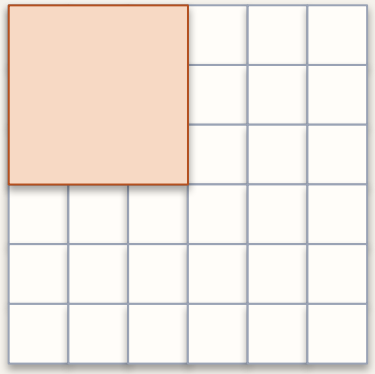
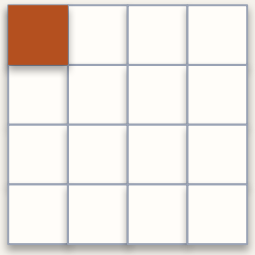


Image d’entrée (6 × 6) et filtre 3 × 3



Carte de caractéristiques (4 × 4)

Localité

Chaque sortie ne regarde qu’une petite zone de pixels voisins.

Partage des poids

Le même filtre (3 × 3 × 3 = 27 poids) sert sur toute l’image.

Un filtre, un motif

Bord vertical, coin, texture : les filtres sont appris, pas écrits à la main.

1 · L'ARCHITECTURE D'UN CNN

On empile des blocs, puis on décide



Les blocs convolution + pooling se répètent plusieurs fois : l'image rétrécit, le nombre de filtres augmente.

Entraînement : perte d'entropie croisée, rétropropagation, descente de gradient (SGD ou Adam).

Plus on va en profondeur, plus les motifs sont abstraits

PREMIÈRES COUCHES

Bords et couleurs

Lignes droites, contrastes, zones claires ou sombres.

COUCHES DU MILIEU

Textures

Carrelage, parquet, tissu, feuillage.

COUCHES PROFONDES

Objets

Robinet, lit, plaque de cuisson, fenêtre.

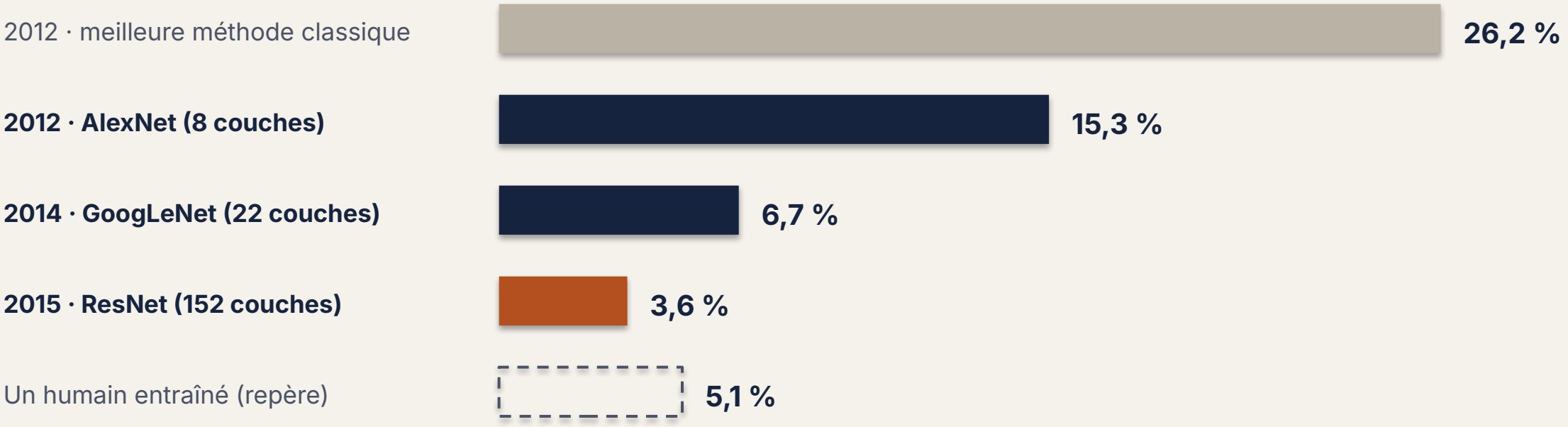
SORTIE

La pièce

« Lit + fenêtre + parquet » : une chambre.

2 · IMAGENET : ERREUR TOP-5, PLUS BAS = MIEUX

En trois ans, l'erreur a été divisée par quatre



Sources : Krizhevsky et al. 2012 · Szegedy et al. 2015 · He et al. 2016 · Russakovsky et al. 2015 · 8 / 17

2 · LE TRANSFER LEARNING

On garde ce que le réseau sait voir, on change seulement la décision

AVANT	**Couches de convolution** entraînées sur ImageNet : 1,2 million d'images	**Tête : 1 000 classes** chien, voiture, guitare…
APRÈS	**Les mêmes couches, gelées** leurs poids ne bougent plus	**Nouvelle tête : 7 pièces** seule partie entraînée

Les premières couches sont générales, les dernières sont spécialisées : c'est ce qui rend le transfert possible (Yosinski et al., 2014).

Source : Yosinski et al., « How transferable are features in deep neural networks? », NeurIPS 2014 · 9 / 17

3 · UN TRAVAIL PUBLIÉ SUR CE SUJET

Les photos aident même à estimer le prix d’un bien

Poursaeed, Matera et Belongie (2018) ont constitué une base de **140 000 photos** classées en **7 types de pièces**.

Un CNN reconnaît la pièce, un second note son niveau de standing. Ces informations sont ajoutées aux données de l’annonce pour estimer le prix de vente.

ERREUR MÉDIANE SUR LE PRIX

Données de l’annonce seules	8,0 %
Zestimate (Zillow)	7,9 %
Avec l’analyse des photos	**5,8 %**

Source : Poursaeed, Matera & Belongie, « Vision-based Real Estate Price Estimation », Machine Vision and Applications, 2018 · 10 / 17

Deux modèles, les mêmes photos, le même test

Données

535 maisons, 2 140 photos, 4 classes. Découpage par maison : 1 496 photos pour l'entraînement, 320 pour la validation, 324 pour le test.

A · CNN depuis zéro

4 couches de convolution, 241 348 poids, tous appris sur nos photos. Augmentation : miroir, zoom, contraste.

B · Transfer learning

ResNet-50 pré-entraîné sur ImageNet, gelé (23,6 M de poids). Seule la nouvelle tête est apprise : 8 196 poids.

Entraînement : entropie croisée, Adam (taux 0,001), arrêt précoce sur la validation.

Outils : Python, TensorFlow / Keras, sur un simple processeur, sans GPU.

De vraies photos d’annonces, prises telles quelles

Salle de bain



Chambre



Façade



Cuisine



Éclairage, cadrage et qualité varient d’une maison à l’autre : c’est ce qui rend la tâche réaliste.

3 · RÉSULTATS SUR LES 324 PHOTOS DE TEST

Le transfer learning gagne nettement

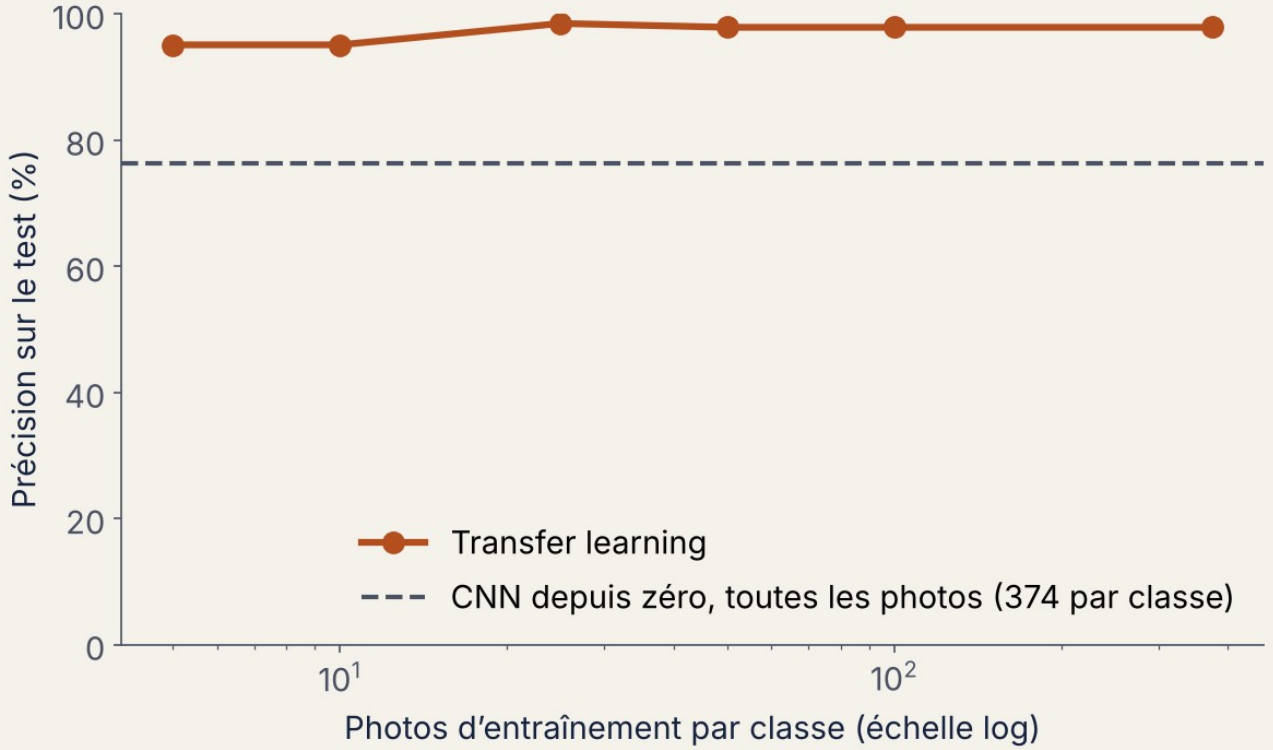
A · CNN depuis zéro

76 %

B · Transfer learning

98 %

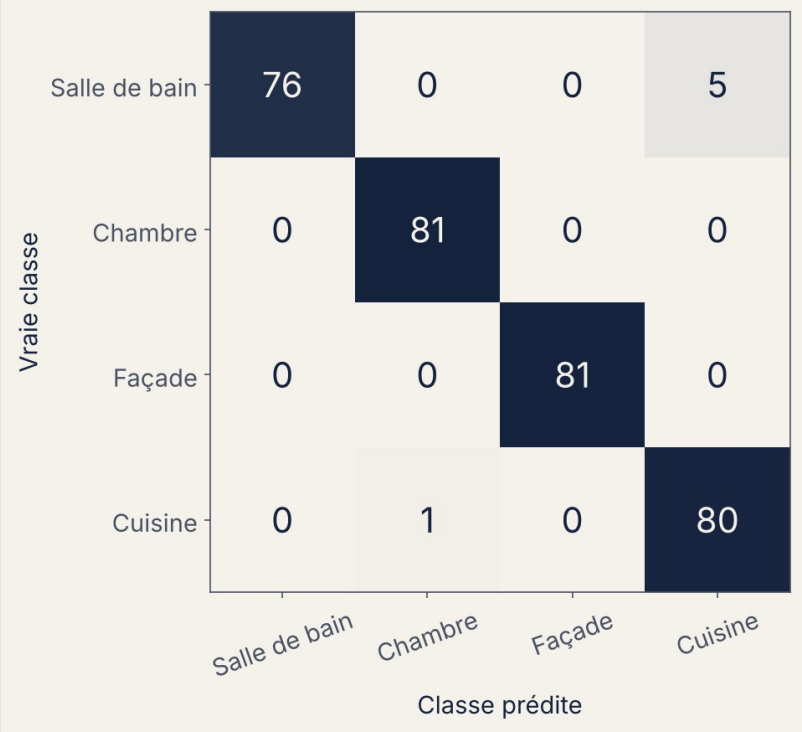
B apprend 30 fois moins de poids et s'entraîne en 56 secondes, contre 4 min 39 pour A.



Précision = part des photos de test bien classées · 13 / 17

3 · OÙ LE MODÈLE SE TROMPE

6 erreurs sur 324, presque toutes au même endroit





5 salles de bain prises pour des cuisines : mêmes meubles en bois, même plan de travail, même carrelage.

1 cuisine prise pour une chambre : une pièce ouverte, avec un canapé au premier plan.

Score F1 par classe : façade 1,00 · chambre 0,99 · salle de bain 0,97 · cuisine 0,96 · 14 / 17

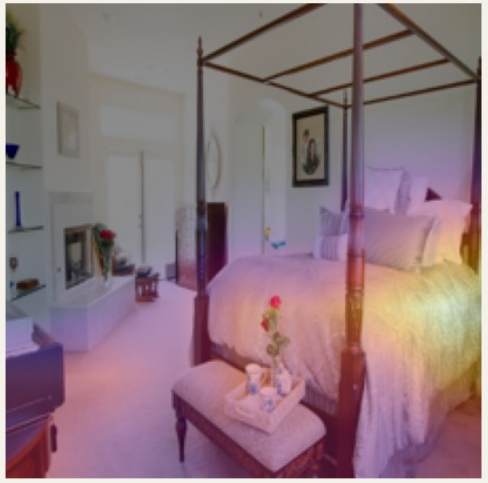
3 · OÙ REGARDE LE RÉSEAU ?

Grad-CAM : les zones qui ont pesé dans la décision

Salle de bain



Chambre



Façade



Cuisine



Les zones claires sont celles qui ont le plus compté. Le réseau s’appuie sur des objets qui ont du sens : le lavabo, le lit, la maison, le plan de travail.

4 · LIMITES

Ce que le modèle ne fera pas tout seul

Pièces ambiguës

Une cuisine ouverte sur le salon appartient aux deux classes. Une seule étiquette ne suffit pas toujours.

Biais des données

Des photos de magazine de décoration ne ressemblent pas aux photos prises au téléphone par un vendeur.

Boîte noire

Grad-CAM montre où regarde le réseau, mais n'explique pas tout son raisonnement.

Portée de mon test

4 classes au lieu de 7, une seule base, un seul découpage. À répéter, et à comparer au fine-tuning.

CONCLUSION

Trois idées à retenir

1

La convolution respecte la structure de l'image : localité et partage des poids.

2

Un CNN apprend une hiérarchie de motifs, des bords jusqu'aux objets.

3

Le transfer learning rend ce savoir réutilisable : 98 % contre 76 % dans mon test.

Merci pour votre attention.