

# IronNet : détection de fers apparents au sein de structures en béton armé

Guillaume Decor<sup>1,2</sup>

Philippe Foucher<sup>1</sup>

Pierre Charbonnier<sup>1</sup>

Fabrice Heitz<sup>2</sup>

<sup>1</sup> Équipe-projet ENDSUM, Cerema, Strasbourg

<sup>2</sup> ICube, UMR 7357, Université de Strasbourg, CNRS

guillaume.decor@cerema.fr

## 1 Introduction

On propose, dans cet article, une démarche exploratoire pour segmenter les armatures métalliques mises à nu sur des structures en béton, à partir d'images photographiques non calibrées et au moyen de méthodes d'apprentissage profond. Ces désordres, appelés *fers apparents*, sont souvent liés à des phénomènes de corrosion des armatures du béton. Les détecter automatiquement et précisément peut apporter une aide à l'expert pour poser un diagnostic sur l'état de la structure. Dans cette contribution, nous proposons un modèle de segmentation des fers apparents basé sur l'architecture U-Net [5] que nous évaluons sur un ensemble d'images d'infrastructures de génie civil.

La plupart des travaux visant à la détection automatique des défauts au sein des ouvrages d'art se concentrent sur la détection de fissures. On peut toutefois noter, ces dernières années, quelques articles relatifs à la détection des fers apparents. Dans [6], une architecture de type Fully Convolutional Networks (FCN) a été employée pour segmenter les fers apparents sur des tabliers de ponts. Dans ce travail, les fers, ainsi que les pertes de matière les entourant, sont répartis dans deux classes distinctes. Une exactitude de 89,7% est atteinte pour les pertes de matière, et de 78,4% pour les fers. Les auteurs de [2] proposent d'appliquer une architecture Faster R-CNN sur des images RGB-D images pour la détection de fers apparents. Sur un jeu de données de 1091 images (apprentissage : 900, test : 191), cette méthode permet d'atteindre une précision moyenne de 90,79%. Une méthode de *meta-learning* est proposée dans [4] afin d'entraîner un classifieur à déterminer la présence d'un défaut ou non dans une image. Plusieurs classes de défauts (fissures, épaufrures, efflorescence, fers apparents, et rouille) sont représentées sur 6481 images d'apprentissage, 616 images de validation et 632 images de test. Entraîné dans un contexte multi-label (i.e. chaque image peut se voir attribuer plusieurs labels si elle présente plusieurs défauts de natures différentes), les auteurs sont parvenus à une exactitude de 70,45% sur le jeu de test. Notons que cette base est disponible pour la communauté scientifique du domaine. Il existe peu de travaux proposant une méthode de segmentation appliquée à ce type de désordres, et nous proposons une nouvelle contribution en ce sens.

## 2 Données

Trois sources de données ont été utilisées dans cette expérimentation : les images des entrées d'un tunnel piéton et de la façade d'un bâtiment universitaire, acquises par nos soins, ainsi que les images de la base CODEBRIM [4] contenant des désordres classés dans la catégorie fers apparents. Les images collectées dans le tunnel piéton ont servi pour l'apprentissage et la validation tandis que les images des deux autres jeux ont été réservées aux tests. Pour renforcer l'indépendance entre les jeux de données, nous avons enregistré les images des deux infrastructures avec des caméras différentes (Sony DSC-HX20V pour le tunnel et Apple iPhone XS ainsi que Samsung SM-T700 pour le bâtiment universitaire).

Le tunnel piéton présente une trentaine de fers apparents situés sur les flancs de ses deux entrées. Les acquisitions ont été réalisées deux fois : une fois par temps ensoleillé, induisant des ombres projetées et une forte variation d'intensité lumineuse, et une autre fois par temps nuageux, l'éclairage étant alors plus homogène. Au total, 212 photographies de résolution  $4896 \times 3672$  ont été prises (134 pour l'apprentissage et 78 pour la validation). Le bâtiment universitaire compte une centaine de fers apparents localisés, pour la plupart, au niveau du quatrième étage. En conséquence, la plupart des photographies ont été prises en contre-plongée. De plus, il est à noter que la distance entre l'objectif et la paroi tend à être bien plus longue dans cette configuration que pour le tunnel piéton (environ 10 m contre 0,1–1 m pour le tunnel). Les défauts apparaissent donc plus petits à l'image. Pour ce jeu de test, 17 images ont été prises par la caméra Samsung ( $2048 \times 1536$ ) et 56 par l'iPhone ( $4032 \times 3024$ ). Ces ensembles seront respectivement notés A et S. La figure 1 présente quelques exemples de ces données et les masques binaires associés, correspondant à l'annotation manuelle effectuée par un opérateur.

La base CODEBRIM portant sur la classification, il a été nécessaire d'établir une vérité terrain au pixel près pour pouvoir tester notre modèle de segmentation. Comme la base décrit plusieurs types de défauts, une première sélection a été effectuée



FIGURE 1 – Extrait du jeu d’apprentissage. **Première ligne** : images originales. **Seconde ligne** : Vérités terrain associées (un pixel blanc dans le masque indique un fer apparent dans l’image correspondante)

	Exactitude	Exactitude pondérée	mIoU	Précision	Rappel	$F_1$ -score
S	93,43	75,08	49,29	5,39	56,50	9,84
A	80,94	85,26	43,47	6,28	89,71	11,75
CODEBRIM	80,49	74,62	60,64	60,03	63,00	61,48

TABLE 1 – Résultats détaillés (en %) obtenus sur les trois jeux de test

pour ne garder que les images présentant des fers apparents. Des 632 images initialement présentes dans le jeu de test, seules 29 ont été conservées. Ensuite, les images ont été segmentées manuellement pour délimiter l’emplacement des fers apparents (en incluant les pertes de matière les entourant).

### 3 Méthodologie et résultats

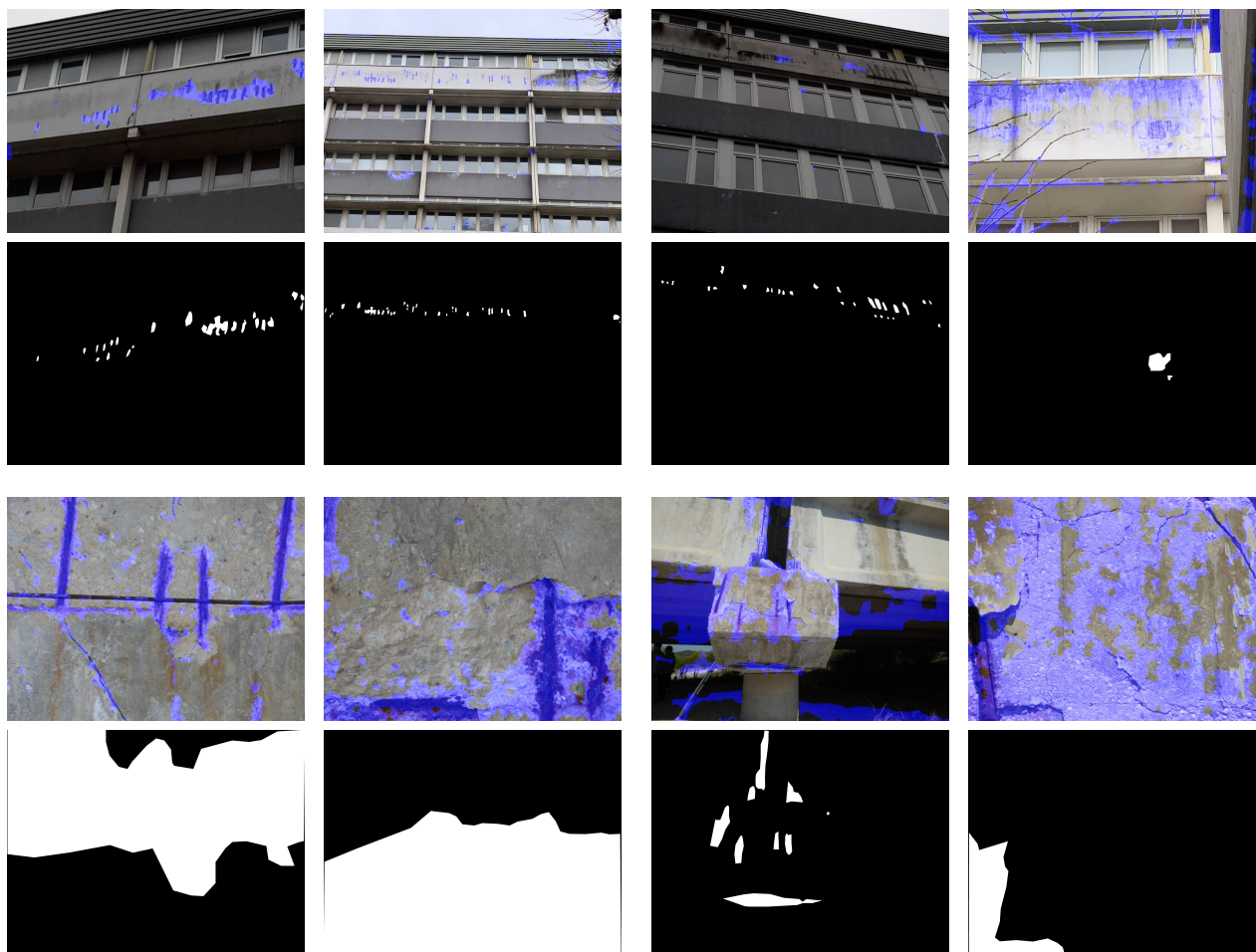
Nous nous positionnons dans le cas d’une segmentation à deux classes : pixels considérés comme fers apparents ou non. L’apprentissage a été réalisé sur un modèle U-Net avec Adam comme optimiseur pendant 1381 époques et l’entropie croisée binaire comme fonction de coût. Ce coût est pondéré par la fréquence d’apparition des différentes classes (saine et fer apparent) de sorte que les erreurs commises sur les zones saines (qui sont majoritaires) soient moins pénalisées que celles portant sur les fers apparents. À chaque époque, les images sont présentées à 4 échelles différentes ( $\times 1$ ,  $\times 1/2$ ,  $\times 1/4$  et  $\times 1/8$ ). Chacune de ces images nouvellement dimensionnées est à son tour dupliquée pour être présente en version originale et en version retournée selon l’axe horizontal. On découpe ensuite une imagerie de taille  $256 \times 256$  par image. D’une seule image, on obtient donc 8 imageries pour l’apprentissage. Celles-ci sont sélectionnées au hasard, si bien que la composition du jeu d’apprentissage varie au fil des époques. L’objectif de cette méthode d’apprentissage est de réduire le temps de calcul nécessaire par époque.

Parmi les modèles générés pendant la phase d’apprentissage, nous retenons celui présentant le meilleur  $F_1$ -score (moyenne obtenue suivant les 4 échelles utilisées) sur le jeu de validation, c’est-à-dire le meilleur compromis entre précision et rappel. Selon ce procédé, c’est le modèle 679 qui est retenu, avec un  $F_1$ -score moyen de 35,53% et une exactitude de 91,73%.

Le test est effectué sur les images en pleine résolution. Les résultats quantitatifs sont reportés dans la table 1 tandis que la figure 2 présente quelques exemples de bonnes détections ainsi que des cas moins bien traités par le modèle. On peut noter que les résultats en termes de précision et de  $F_1$ -score sont assez faibles, même si le rappel et les exactitudes (brute et pondérée) demeurent corrects. Ceci s’explique par un niveau élevé de vrai négatif (i.e. pixels classés à raison comme sains). Notons, par ailleurs, que les performances en rappel et précision obtenues sur la base CODEBRIM sont plus homogènes entre elles. Cependant, comme le montrent les images de la figure 2, les défauts de la base CODEBRIM apparaissent en gros plan et le déséquilibre d’effectifs entre les classes est beaucoup plus faible que pour les jeux de test A et S.

### 4 Conclusion et perspectives

Bien que les résultats laissent transparaître le potentiel de cette méthode, il reste une marge de progression significative. Plusieurs pistes peuvent être proposées pour poursuivre ce travail exploratoire. Tout d’abord, le modèle utilisé ici est assez simple et de nombreux modèles plus sophistiqués dédiés à la segmentation sémantique ont émergé depuis [1] et seront



(a) Bonnes détections

(b) Cas difficiles

FIGURE 2 – Exemples de détections sur les jeux de test. **Rangées du haut** : résultats sur les images du bâtiment universitaire (A ou S). **Rangées du bas** : résultats sur les images de la base CODEBRIM. Les zones surlignées en bleu correspondent aux pixels classés comme fers apparents et les vérités terrain sont présentées en dessous de chaque prédiction.

prochainement mis en œuvre. Il serait par ailleurs intéressant de mesurer l'impact de méthodes d'adaptation de domaine, à l'image de celle proposée par Ganin et al. [3], sur ce type de données mettant en jeu des acquisitions se faisant souvent dans des conditions non contrôlées, avec une variabilité importante.

## Références

- [1] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12) :2481–2495, 2017.
- [2] G. H. Beckman, D. Polyzois, and Y.-J. Cha. Deep learning-based automatic volumetric damage quantification using depth camera. *Automation in Construction*, pages 114–124, 2019.
- [3] Y. Ganin and V. Lempitsky. Unsupervised Domain Adaptation by Backpropagation. In *International Conference on International Conference on Machine Learning (ICML)*, volume 37, pages 1180–1189, Lille, France, 2015.
- [4] M. Mundt, S. Majumder, S. Murali, P. Panetsos, and V. Ramesh. Meta-Learning Convolutional Neural Architectures for Multi-Target Concrete Defect Classification With the COncrete DEfect BRidge IMage Dataset. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11196–11205, Long Beach, California, USA, 2019.
- [5] O. Ronneberger, P. Fischer, and T. Brox. U-Net : Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Munich, Germany, 2015.
- [6] J. J. Rubio, T. Kashiwa, T. Laiteerapong, W. Deng, K. Nagai, S. Escalera, K. Nakayama, Y. Matsuo, and H. Prendinger. Multi-class structural damage segmentation using fully convolutional networks. *Computers in Industry*, 112 :103–121, 2019.