

Problématique

Chaque année, plus de 30 millions d'images de formulaires fédéraux sont numérisées à Matane afin de faciliter la consultation et la préservation. Cependant, dû à la nature de plusieurs de ces documents, une retranscription manuelle est nécessaire pour permettre le passage au format numérique.

Objectifs

- Automatiser le processus de retranscription
- Permettre un taux de reconnaissance de plus de 85% sur des documents structurés (ex: formulaires avec des cases délimitées)
- Permettre un taux de reconnaissance de 75% sur des documents n'ayant pas une structure définie (ex: documents de nature historique)
- Traiter plus d'un document à la minute sans interruption (24/7)

Méthodologie

- Prétraitement de l'image et corrections
- Utilisation des avancées récentes en intelligence artificielle
- Identification des cas complexes pour révision

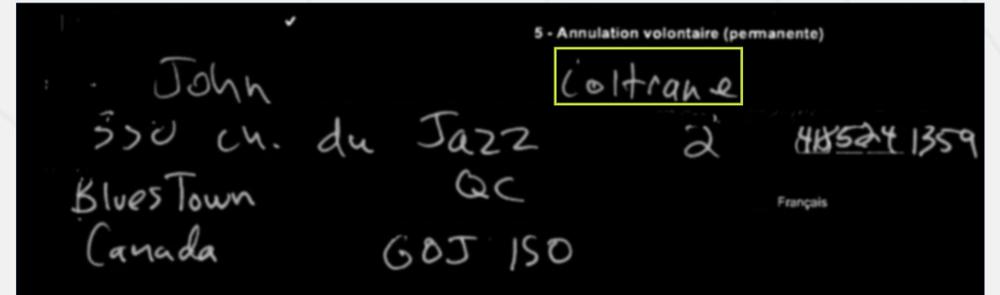


Figure 3: Extraction du texte et segmentation

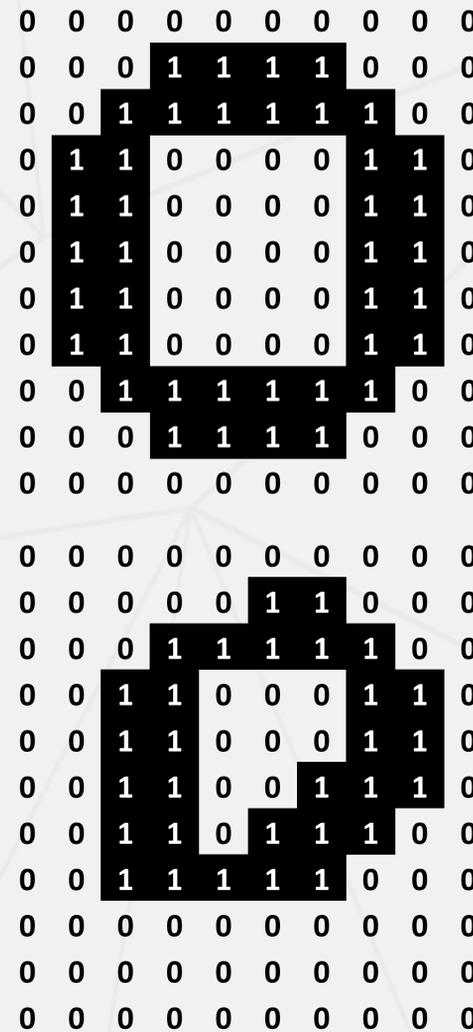


Figure 2: Représentation d'une image d'un zéro idéal et d'un zéro réaliste

Poursuite du projet

- Généralisation à l'ensemble des documents et formulaires fédéraux
- Gestion de l'écriture cursive
- Intégration aux services présentement offerts à la Direction de l'imagerie et opérations du Receveur Général à Matane
- Apprentissage et amélioration du procédé de manière continue et automatique

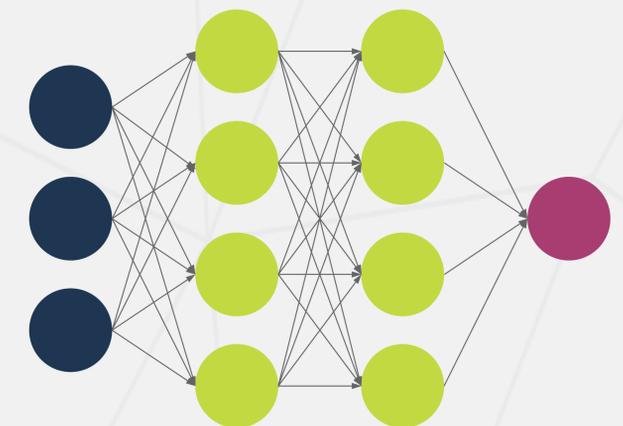


Figure 4: Schématisation d'un réseau de neurones à connectivité totale

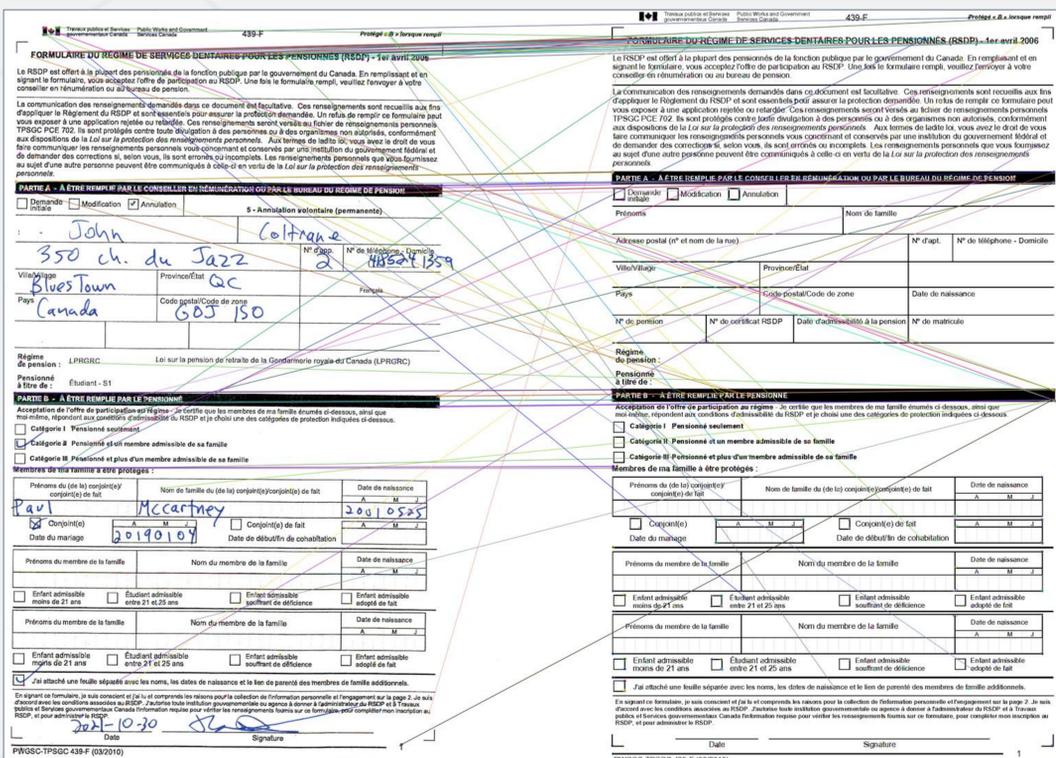


Figure 1: Ajustement du formulaire et correction de la numérisation