

RECONNAISSANCE OPTIQUE DE LA MUSIQUE BASÉE SUR L'APPRENTISSAGE MACHINE À GRANDE ÉCHELLE POUR DES DONNÉES DE PARTITIONS MUSICALES

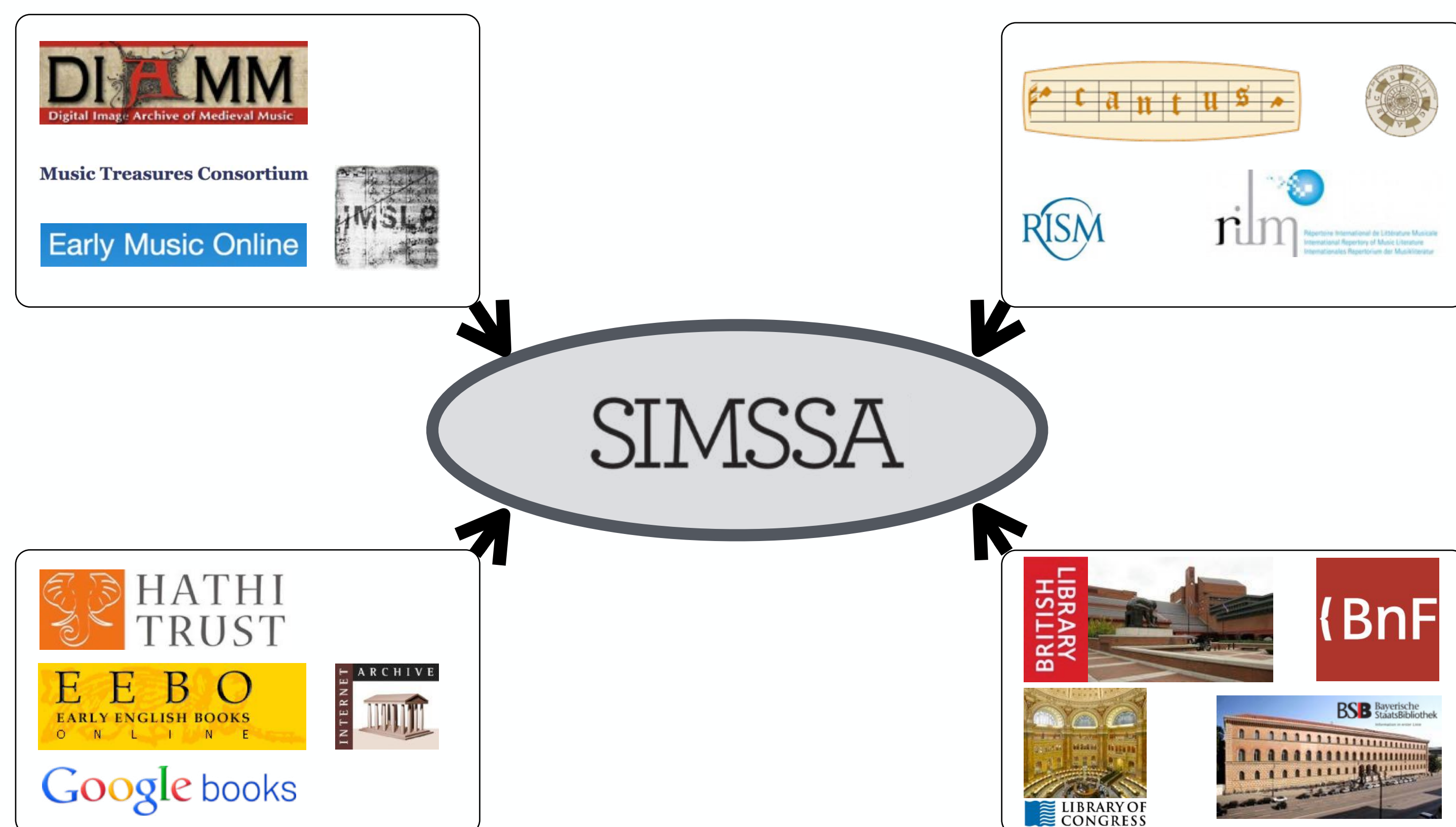
Gabriel Vigliensoni
gabriel@music.mcgill.ca

Jorge Calvo-Zaragoza
jcalvo@dlsi.ua.es

Ichiro Fujinaga
ich@music.mcgill.ca

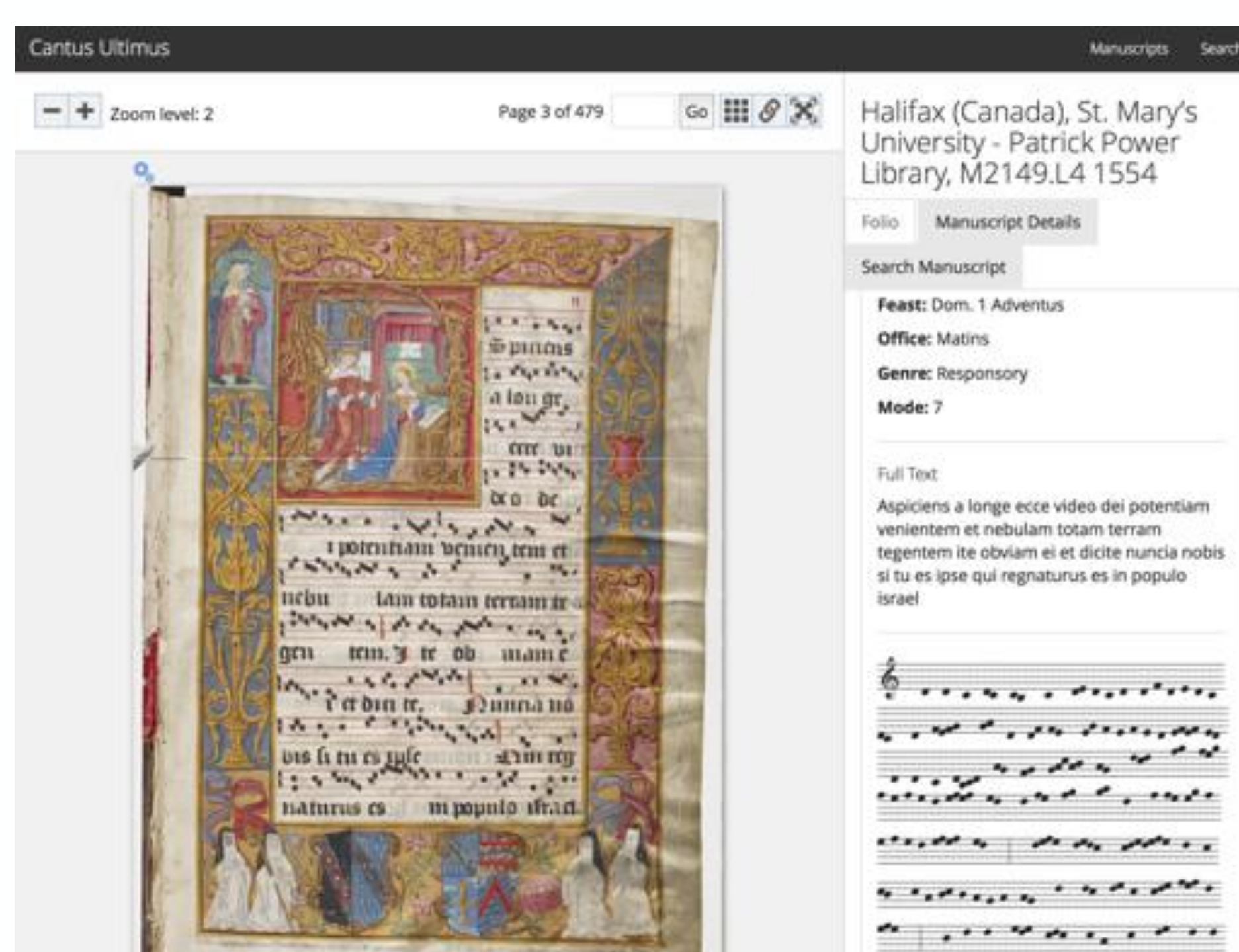
Notre vision

- À travers le projet Single Interface for Music Score Searching and Analysis (SIMSSA), nous cherchons à bâtir une infrastructure capable d'analyser et de chercher des millions d'images de pages musicales.



Objectifs de l'étude

- Créer un site internet afin de rechercher des partitions dans les librairies musicales, à la « Google scores » mais sans utiliser les outils de Google.
- Créer des outils pour convertir des partitions numérisées dans un format symbolique pouvant être recherchées en utilisant la reconnaissance musicale optique (ROM).
- Développer des interfaces pour voir, rechercher et analyser des partitions à partir de plusieurs sources différentes.

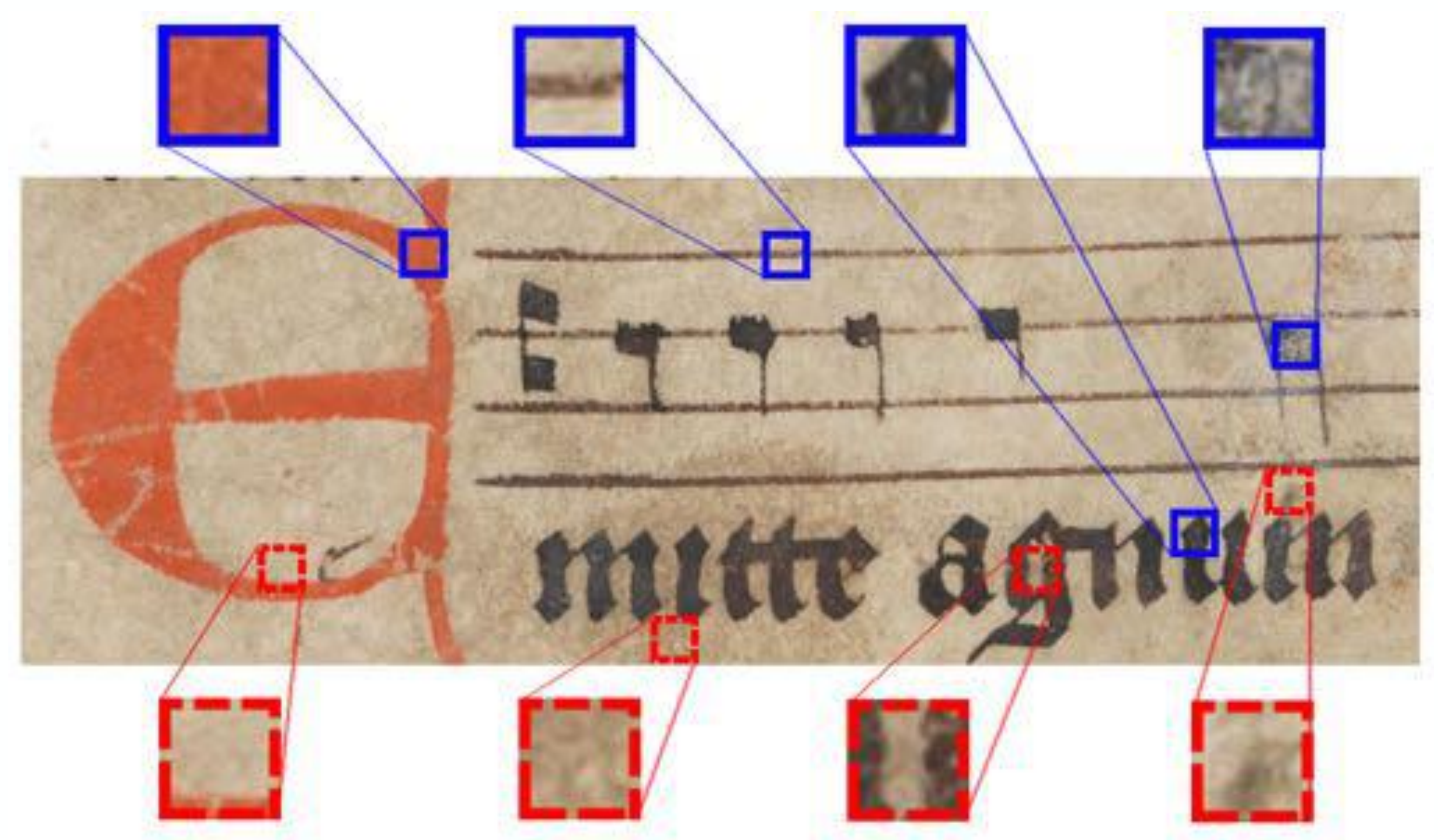


Prototype actuel : Cantus Ultimus (<https://cantus.simssa.ca/>)

- La ROM est une technologie équivalente à la reconnaissance optique de caractères, mais appliquée au contexte de la musique.
- Les méthodes traditionnelles de ROM utilisent des stratégies heuristiques pour reconnaître les différents éléments dans des partitions musicales.
- Ces méthodes dépendent beaucoup des caractéristiques uniques des images et il est donc difficile de les appliquer à de grands volumes de documents.

Cadre de l'étude

- Avec le projet SIMSSA, nous développons des systèmes de ROM basés sur l'infonuagique, afin d'optimiser le processus de reconnaissance de documents musicaux en permettant la correction participative et l'apprentissage machine distribué.
- Ces approches novatrices à la ROM requièrent des systèmes de stockage à grande échelle et des grappes de serveurs.
- Nous présentons une approche novatrice à la ROM qui utilise l'apprentissage machine pour classer les pixels des partitions musicales dans leurs catégories potentielles.
- Cette méthode assume que la zone entourant chaque pixel contient assez d'information pour pouvoir discriminer leur catégorie.



- Le pixel sera étiqueté en fonction de ses informations environnante.

$$l(x, y) \rightarrow \{ \text{contexte,} \\ \text{lignes de la portée,} \\ \text{symboles musicaux,} \\ \text{paroles, ...} \}$$

- Cette méthode utilise ces données comme caractéristiques pour entraîner un réseau neuronal convolutif.

Conclusion

- Notre système de ROM démontre une performance comparable aux algorithmes de pointe du domaine.
- Mais il présente plus d'extensibilité et d'évolutivité parce qu'il a seulement besoin de plus de données pour apprendre un meilleur modèle.

Remerciements

- Ce travail a été supporté par le Fond de recherche du Québec-Société et culture, le Conseil de recherches en sciences humaines du Canada, et le Ministerio español de Educación, Cultura y Deporte par le biais d'une bourse FPU (Réf. AP2012--0939).