



AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mr **EL ASRI Mohamed**
Soutiendra : le Samedi 16/05/2026 à 10H00
Lieu : **FSDM – Département d'Informatique**

Une thèse intitulée :

Deep Generative Models for Sketch-Based Image Generation and Blind Image Inpainting

En vue d'obtenir le **Doctorat**

FD : Sciences et Technologies de l'Information et de la Communication
Spécialité : Informatique

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
YAHYAOUY Ali	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
BELLACH Benaissa	Ecole Nationale des Sciences Appliquées, Oujda	PES	Rapporteur
TAIME Abderazzak	École Supérieure de Technologie, Khénifra	MCH	Rapporteur
RIFFI Jamal	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Rapporteur
MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Examineur
RAMADAN Hiba	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Examineur
ELHARROUSS Omar	Université des Émirats Arabes Unis	MC	Expert
TAIRI Hamid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse



Résumé :

Les avancées récentes en apprentissage profond ont profondément transformé le domaine de la vision par ordinateur, en particulier dans les tâches de génération et de restauration d'images. Les modèles génératifs ont démontré des capacités remarquables pour synthétiser des images visuellement réalistes à partir de différentes formes de données d'entrée. Toutefois, produire des images qui préservent simultanément la précision structurelle, le réalisme des textures et la cohérence sémantique globale demeure un défi important, notamment dans les situations impliquant des entrées visuelles incomplètes ou des informations structurelles limitées. Cette thèse étudie le problème de la génération d'images à partir de croquis réalisés à main levée ainsi que la restauration de contenus visuels dégradés à travers la tâche d'inpainting d'images. L'objectif principal est de développer des cadres génératifs profonds capables de modéliser efficacement à la fois les détails visuels locaux et les dépendances structurelles à longue portée. Pour atteindre cet objectif, cette thèse explore les avancées récentes dans les réseaux antagonistes génératifs (GAN), les architectures basées sur les Transformers, ainsi que les représentations multi-résolution. Dans un premier temps, une revue complète des méthodes existantes de génération d'images est présentée. Elle couvre plusieurs paradigmes, notamment la traduction image-à-image, la génération d'images guidée par croquis, la génération conditionnelle et les approches génératives basées sur les Transformers. Cette analyse met en évidence l'évolution des modèles génératifs profonds ainsi que les principaux défis liés à la préservation de la structure et au raisonnement contextuel. Dans un second temps, un nouveau cadre de génération conditionnelle d'images, nommé S2TRGAN, est proposé pour traduire des croquis réalisés à main levée en images visuellement cohérentes. L'architecture proposée combine l'extraction de caractéristiques convolutionnelles avec un module Vision Transformer intégré dans un générateur de type U-Net, permettant ainsi de capturer simultanément les textures fines et les relations géométriques à longue distance. Afin de soutenir cette tâche, un nouveau jeu de données intitulé Sketch2Tile est introduit. Il contient des paires de croquis dessinés à la main et d'images haute résolution de carreaux de zellige marocain. Les résultats expérimentaux montrent que la méthode proposée surpasse plusieurs approches existantes de traduction image-à-image en termes de cohérence structurelle et de réalisme visuel. Enfin, cette thèse aborde le problème de l'inpainting d'images en mode aveugle, dans lequel les régions corrompues ne sont pas connues à l'avance et doivent être détectées et reconstruites automatiquement. Une architecture générative hybride combinant extraction convolutionnelle, décomposition multi-résolution basée sur les ondelettes et modélisation contextuelle à l'aide de Transformers est proposée afin d'améliorer la reconstruction de structures complexes et de textures détaillées. Des expérimentations approfondies menées sur des jeux de données de référence tels que CelebA-HQ et FFHQ démontrent que l'approche proposée atteint des performances compétitives par rapport aux méthodes de l'état de l'art. Dans l'ensemble, les contributions de cette thèse participent à l'avancement des modèles génératifs profonds pour la synthèse et la restauration d'images structurées. En intégrant guidage structurel, représentations multi-échelles et raisonnement contextuel global, les méthodes proposées offrent des solutions plus robustes et plus précises pour les tâches de génération d'images à partir de croquis et d'inpainting d'images.

Mots clés :

Modèles profonds, réseaux antagonistes génératifs (GAN), Transformers, Inpainting d'images en mode aveugle, Traduction image-à-image, Traduction croquis-à-image.



DEEP GENERATIVE MODELS FOR SKETCH-BASED IMAGE GENERATION AND BLIND IMAGE INPAINTING

Abstract:

Recent advances in deep learning have significantly transformed the field of computer vision, particularly in the domain of image generation and restoration. Generative models have demonstrated remarkable capabilities in synthesizing visually realistic images from various forms of input data. However, producing images that simultaneously preserve structural accuracy, texture realism, and global semantic coherence remains a challenging task, especially in scenarios involving sparse inputs or incomplete visual information. This thesis investigates the problem of image generation from freehand sketches and the restoration of degraded visual content through image inpainting. The primary objective is to develop deep generative frameworks capable of effectively modeling both local visual details and long-range structural dependencies. To achieve this goal, the thesis explores recent advances in generative adversarial networks, transformer based architectures, and multi-resolution feature representations. First, a comprehensive review of existing image generation methodologies is conducted, covering multiple paradigms including image-to-image translation, sketch-to-image synthesis, conditional image generation, and transformer-based generative modeling. This review analyzes the evolution of deep generative models and high lights key challenges related to structural preservation and contextual reasoning. Second, a novel conditional image generation framework, termed S2TRGAN, is proposed for translating freehand sketches into visually coherent images. The proposed architecture integrates convolutional feature extraction with a Vision Transformer bottleneck within a U-Net generator, enabling the model to capture both fine-grained textures and long-range geometric relationships. To support this task, a new dataset named Sketch2Tile is introduced, consisting of paired freehand sketches and high-resolution images of Moroccan zellige tiles. Experimental results demonstrate that the proposed method achieves superior performance compared to existing image-to-image translation approaches in terms of structural consistency and visual realism.

Regions are unknown and must be implicitly detected and reconstructed. A hybrid generative architecture combining convolutional feature extraction, wavelet-based multi-resolution decomposition, and transformer-based contextual modeling is proposed to improve the reconstruction of complex structures and textures. Extensive experiments conducted on benchmark datasets such as CelebA-HQ and FFHQ demonstrate that the proposed approach achieves competitive performance compared to state-of-the-art methods. Overall, the contributions of this thesis advance the understanding and development of deep generative models for structured image synthesis and restoration. By integrating structural guidance, multi-scale feature representations, and global contextual reasoning, the proposed frameworks provide more robust and accurate solutions for sketch-guided image generation and image inpainting tasks.

Key Words:

Deep Models, Generative adversarial networks (GAN), Transformers, Blind image inpainting, Image-to-image translation, Sketch-to-image translation.