New Approach for Mapping Land Cover from Archive Grayscale Satellite Imagery

Mohamed Rabii Simou, Mohamed Maanan, Safia Loulad, Mehdi Maanan, Hassan Rhinane

technologies

MDPI

Article

New Approach for Mapping Land Cover from Archive Grayscale Satellite Imagery

Mohamed Rabii Simou [1,2], Mohamed Maanan [1,*], Safia Loulad [2], Mehdi Maanan [2] and Hassan Rhinane [2]

[1] UMR 6554 CNRS LETG-Nantes Laboratory, Institute of Geography and Planning, Nantes University, 44312 Nantes, France; mohamed-rabii.simou@etu.univ-nantes.fr
[2] GEOPEN Laboratory, Earth Sciences Department, Faculty of Sciences-Ain Chock, University Hassan II, Casablanca 20000, Morocco; s.loulad1@gmail.com (S.L.); mehdi.maanan@gmail.com (M.M.); h.rhinane@gmail.com (H.R.)
* Correspondence: mohamed.maanan@univ-nantes.fr

Abstract: This paper examines the use of image-to-image translation models to colorize grayscale satellite images for improved built-up segmentation of Agadir, Morocco, in 1967 and Les Sables-d'Olonne, France, in 1975. The proposed method applies advanced colorization techniques to historical remote sensing data, enhancing the segmentation process compared to using the original grayscale images. In this study, spatial data such as Landsat 5TM satellite images and declassified satellite images were collected and prepared for analysis. The models were trained and validated using Landsat 5TM RGB images and their corresponding grayscale versions. Once trained, these models were applied to colorize the declassified grayscale satellite images. To train the segmentation models, colorized Landsat images were paired with built-up-area masks, allowing the models to learn the relationship between colorized features and built-up regions. The best-performing segmentation model was then used to segment the colorized declassified images into built-up areas. The results demonstrate that the Attention Pix2Pix model successfully learned to colorize grayscale satellite images accurately, improving the PSNR by up to 27.72 and SSIM by 0.96. Furthermore, the results of segmentation were highly satisfactory, with UNet++ identified as the best-performing model with an mIoU of 96.95% in Greater Agadir and 95.42% in Vendée. These findings indicate that the application of the developed method can achieve accurate and reliable results that can be utilized for future LULC change studies. The innovative approach of the study has significant implications for land planning and management, providing accurate LULC information to inform decisions related to zoning, environmental protection, and disaster management.

Keywords: deep learning; remote sensing; grayscale colorization; built-up segmentation; historical satellite imagery

Academic Editor: Ioannis K. Brilakis

Received: 4 March 2025
Revised: 7 April 2025
Accepted: 10 April 2025
Published: 14 April 2025

Citation: Simou, M.R.; Maanan, M.; Loulad, S.; Maanan, M.; Rhinane, H. New Approach for Mapping Land Cover from Archive Grayscale Satellite Imagery. *Technologies* **2025**, *13*, 158. https://doi.org/10.3390/technologies13040158

1. Introduction

Historical satellite imagery offers a unique window into the past [1]. However, much of these archived data exist only in grayscale, limiting the ability to extract detailed features and perform segmentation [2]. As a result, achieving accurate land use and land cover (LULC) mapping, which is essential to understand urban growth, environmental transformations, and planning sustainable interventions, requires overcoming these limitations.

Early satellite missions have played a pivotal role in shaping remote sensing research [3]. For example, the Landsat program has been operational since 1972, providing multispectral data; since Landsat 4 in 1982, it has offered data at a spatial resolution of

30 m, laying the foundation for decades of environmental monitoring and analysis [4]. Even earlier, the Corona satellite program of the 1960s produced grayscale images with resolutions around 1.8 m. Initially used by US intelligence agencies for reconnaissance and cartography, these declassified images became accessible to the scientific community in 1995 [5]. More recently, the Sentinel-2 mission, launched in 2015, further advanced remote sensing by offering 10 m resolution multispectral imagery, thereby expanding the toolkit available for modern environmental and urban studies [6]. The traditional methods for analyzing these images often relied on manual interpretation or rudimentary digital techniques, which are both time-consuming and prone to errors [2]. The limitations inherent in grayscale imagery pose challenges for the accurate segmentation and classification of land cover, especially when attempting to delineate built-up areas from natural features.

With the advent of deep learning, image-to-image translation techniques have become the leading approach for enhancing historical satellite imagery. Pioneered by the introduction of generative adversarial networks (GANs) in 2014 [7], followed by methods based on conditional GANs (cGANs), also introduced in 2014 [8], these approaches have demonstrated significant potential in translating grayscale images into realistic colorized representations. The subsequent developments include Pix2Pix, developed in 2017 [9], ChromaGAN, presented in 2019 [10], BigColor, proposed in 2022 [11], and iColoriT, introduced in 2023 [12], each addressing the challenge of colorizing images in unique ways. These methods enable the transformation of grayscale images into colorized versions by learning the underlying mapping between the two domains, thereby enhancing the visual and spectral information available for subsequent analyses.

In addition to improved colorization, the segmentation of land use and land cover (LULC) has also benefited from advancements in deep learning. Modern semantic segmentation models, including FPN, introduced in 2017 [13], DeepLabV3+, developed in 2018 [14], UNet++, proposed in 2018 [15], and Segformer, presented in 2021 [16], have been employed to effectively delineate different LULC classes. These models leverage high-level features and spatial context to accurately classify various land cover types, such as built-up areas, vegetation, and water bodies, thereby providing more reliable datasets.

Several studies have contributed to the progress of historical grayscale imagery over the years. Cohn in 2017 [17] aimed to enhance the spectral utility of grayscale satellite imagery by employing generative adversarial networks (GANs) to colorize it, thus improving spectral information for applications like object detection. Following this, Gravey et al. in 2018 [18] sought to improve the spectral resolution of historical satellite data, using multiband spatial pattern matching to transfer spectral characteristics from modern imagery such as Landsat-8 to older datasets such as Corona, preserving spatial and spectral coherence. Also in 2018, Li et al. [19] aimed to simultaneously enhance the resolution and color in archived satellite imagery, developing a multitask deep neural network that performed super-resolution and colorization. Advancing to 2021, Themistocleous et al. [20] aimed to reconstruct historical land cover from grayscale imagery, applying the DeOldify framework, an improved GAN-based method, to colorize a 1963 Corona image over Larnaca, Cyprus, achieving high-quality results, showcasing its utility for historical land cover analysis. Most recently, Remondino et al. in 2022 [21] aimed to develop an effective colorization tool for historical aerial images, introducing Hyper-U-NET, a novel neural network combining U-NET-like architecture with HyperConnections, outperforming other CNN- and GAN-based methods in diverse scenarios like urban and rural landscapes. Shamsaliei et al. in 2024 [22] aimed to establish a benchmark for semantic segmentation of historical aerial imagery, presenting HAIR, a dataset of grayscale images spanning 1947 to 1998, and assessed state-of-the-art semantic segmentation models on it, highlighting

challenges such as grayscale limitations and camera technology variations, thus paving the way to improve segmentation for historical data.

However, despite these advancements, the existing methods have limitations when applied to historical grayscale satellite imagery for built-up-area detection. Segmentation models applied directly to grayscale images often struggle with accuracy due to the lack of color information, which is crucial for distinguishing various land cover types [22]. Meanwhile, image-to-image translation techniques, while effective for colorizing grayscale images [23], have primarily been used for visual enhancement rather than as a preprocessing step for segmentation tasks. This study addresses these gaps by proposing a novel two-step approach: first, colorizing historical grayscale images using state-of-the-art image-to-image translation models and then segmenting these colorized images to accurately detect built-up areas. By integrating these steps, we aim to leverage both enhanced visual quality and improved feature representation from colorized images, particularly for the unique challenges of Agadir, Morocco (1967), and Les Sables-d'Olonne, France (1975).

2. Materials and Methods

2.1. Study Area

Two study areas have been selected for this paper: Agadir, Morocco, and Les Sables-d'Olonne, France, two Atlantic coastal destinations (Figure 1).

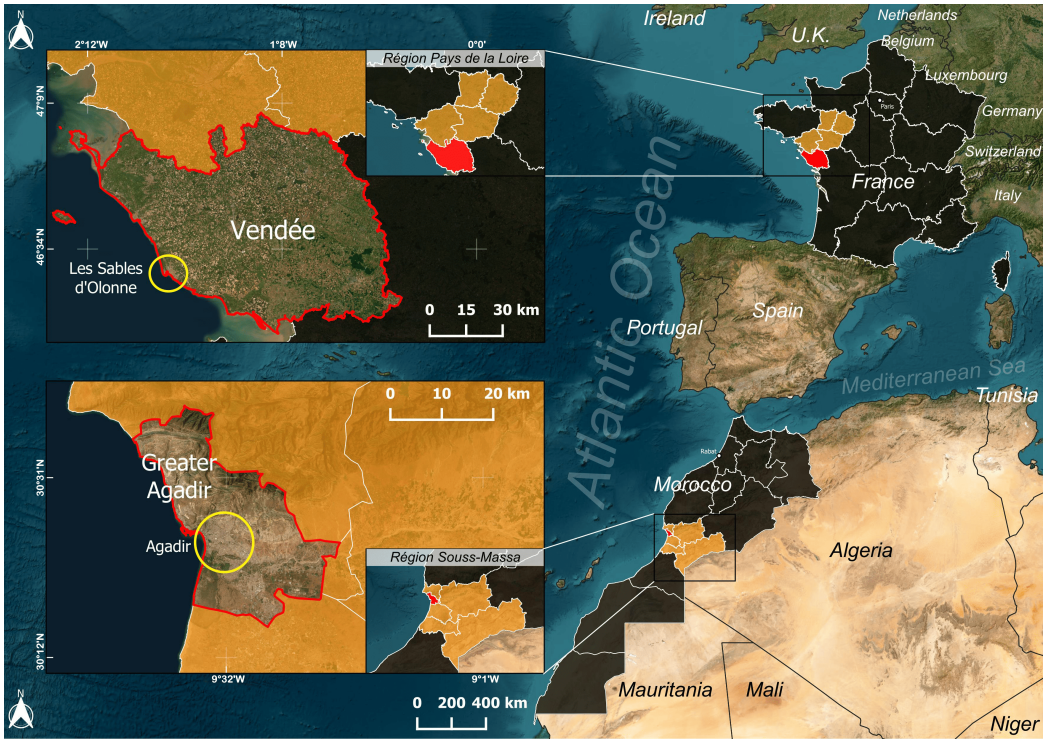

Figure 1. Study area.

Agadir, Morocco, situated in southwestern Morocco, is the capital of the Souss-Massa region. According to the 2024 General Population and Housing Census (RGPH 2024) conducted by the Haut Commissariat au Plan (HCP) [24], the population of the Greater Agadir area is 1,267,000 people. The city of Agadir itself has a population of approximately 501,000 people. Agadir has a temperate climate, characterized by mild temperatures and abundant sunshine. Temperatures typically range from 14 to 16 degrees Celsius in the early months of the year, warming up to 20 to 25 degrees Celsius in July [25]. The city enjoys an average of nearly 300 days of sunshine per year, making it a popular destination for tourists seeking warm, sunny weather [25]. The modern infrastructure of Agadir is largely

due to its remarkable recovery and redevelopment following a devastating earthquake in 1960 [26]. Today, the city thrives on tourism thanks to its long sandy beaches and vibrant cultural attractions, making it a vital economic hub in southern Morocco.

Les Sables-d'Olonne, France, located in the Vendée department of the Pays de la Loire region, is a subprefecture with a population of approximately 48,000 residents (2021 INSEE) [27]. The population of Vendée department is nearly 700,000 (2021 INSEE) [28]. This coastal town lies in the Bay of Biscay and experiences a temperate oceanic climate. Winters are generally cool and damp, while summers remain mild and sunny, with rainfall evenly distributed throughout the year [29]. Famous for its beautiful beaches, Les Sables-d'Olonne has become a major destination for seaside tourism, further highlighted by its status as the starting point for the renowned Vendée Globe yacht race [30], an event that underscores the town's maritime heritage and international appeal.

2.2. Workflow

This study employs a structured workflow to implement the new approach (Figure 2).

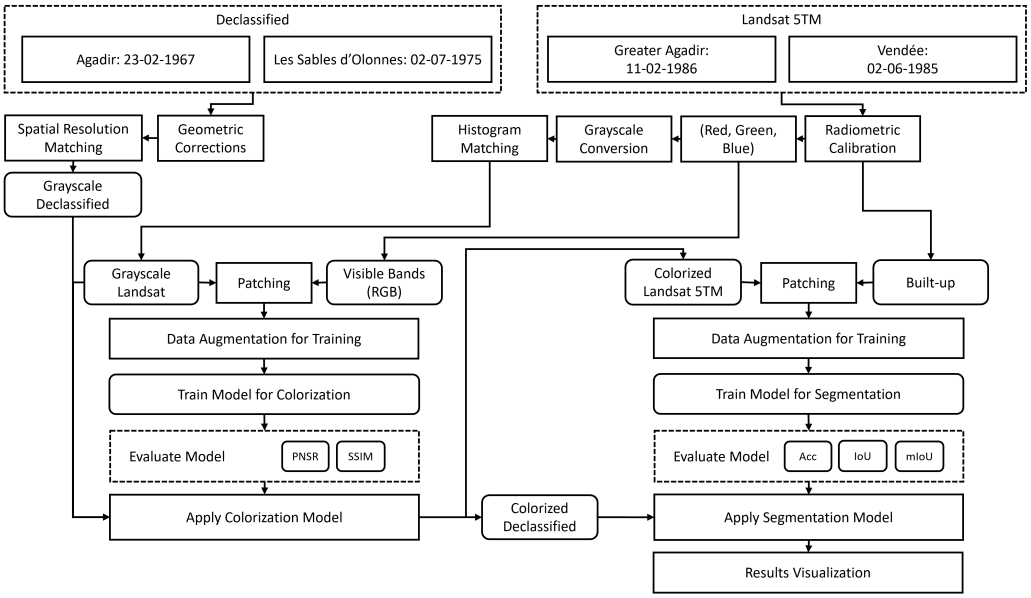

Figure 2. Methodology of the study.

2.3. Data Collection

This phase of the study involved downloading declassified grayscale satellite scenes from the USGS Earth Explorer platform. The Corona image for Agadir was taken on 23 February 1967 by the KH-4B mission; the Entity ID for the scene is DS1039-1011DA002, with a resolution ranging from 1.8 to 2.75 m. The Les Sables-d'Olonne scene was captured on 2 July 1975 by the KH-9 Hexagon mission; the Entity ID for the scene is DZB1210-500097L005001, with a resolution ranging from 6 to 9 m (Table 1).

Table 1. Declassified satellite data for Agadir and Les Sables-d'Olonne.

City	System	Entity ID	Resolution	Date	Source
Agadir	KH-4B	DS1039-1011DA002	1.8 to 2.75 m	23 February 1967	USGS Earth Explorer
Les Sables-d'Olonne	KH-9	DZB1210-500097L005001	6 to 9 m	2 July 1975	USGS Earth Explorer

Additionally, Landsat 5TM satellite images were downloaded from Google Earth Engine (GEE). The Landsat 5TM images were selected because of the availability of visible bands with a resolution of 30 m. The images were captured on 11 February 1986 for Greater Agadir and 2 June 1985 for Vendée. These dates were chosen to capture similar land cover patterns to the corresponding grayscale images while ensuring that the available data were of suitable quality for analysis (Table 2).

Table 2. Landsat 5TM data for Agadir and Les Sables-d'Olonne.

City	Type	Spatial Resolution	Date	Source
Greater Agadir	Landsat 5TM	30 m	11 February 1986	GEE
Vendée	Landsat 5TM	30 m	2 June 1985	GEE

2.4. Data Preparation

Subsequently, a series of preprocessing steps were applied to both the declassified and Landsat 5 satellite images.

For the declassified images, geometric corrections were performed to georeference and correct any positional inaccuracies and ensure proper alignment with the coordinate system. Following that, spatial resolution matching was conducted to downscale declassified images resolutions to match the Landsat images at 30 m. This ensured that both datasets had consistent spatial characteristics. Lastly, the images were saved as grayscale declassified images.

For the Landsat 5 images, radiometric calibration of the bands was conducted to adjust for any sensor-related inconsistencies and transform DN values to TOA reflectance. Then, a manual annotation was conducted on the Landsat dataset to create built-up masks. The visible bands (RGB) were separated into two datasets: one maintaining the original RGB format and another created by grayscale conversion. For the grayscale conversion, a standard luminance formula [31] was applied, using weights (0.299R + 0.587G + 0.114B) to accurately represent the contribution of each channel. Next, histogram matching [32] was performed to ensure that Landsat resembles the grayscale nature of the original declassified image grayscale data.

After finishing preparation, the outputs included the grayscale declassified images, the built-up masks, visible RGB bands, and the grayscale Landsat image.

2.5. Training Models for Colorization

2.5.1. Training Data for Colorization

To generate suitable input–output pairs for training, the data underwent a series of specific preparation steps for the model:

- **Patching**: The grayscale and RGB Landsat images were initially padded to ensure their dimensions were divisible by the specified patch size. This step allowed the images to be uniformly divided into patches of size 256 × 256 pixels, creating manageable and consistent training samples.
- **Data Augmentation**: To increase the diversity of the training dataset, data augmentation techniques were employed. These transformations included rotations, translations, horizontal flips, shear transformations, and zooming.

2.5.2. Generative Adversarial Networks

A generative adversarial network (GAN) is a type of neural network that involves two competing networks: a generative neural network (G) and a discriminative neural network (D) [7]. The G network takes a random noise vector (z) as input and generates an output

vector (y) using the generator function $G : z \rightarrow y$. The goal is for the distribution of the output ($p_z(z)$) to match the distribution of the input (real data) as closely as possible ($p_x(x)$). The D network determines whether its input is real or fake, enabling it to distinguish between whether it originated from $p_x(x)$ or $p_z(z)$. Ultimately, the objective of GANs is to improve the performance of G until D is no longer able to differentiate between real and synthesized images [7]. To achieve this goal, the models are trained through a zero-sum game designed to discover the optimal mapping function, Equation (1).

$$G^* : \arg \min_G \max_D \mathcal{L}_{\text{GAN}}(G, D) \tag{1}$$

$\mathcal{L}_{\text{GAN}}(G, D)$ is the objective function for GANs, involving the real data vector x, the probability distribution of the real data $p_x(x)$, the random noise vector z, and the probability distribution of the noise data $p_z(z)$. The training process of GANs involves alternating between the training of the discriminative neural network (D) and the generative neural network (G), thereby enhancing the quality of both models. It is important to balance the improvement of each model during training to prevent one from failing the other. The backpropagation algorithm is utilized for training. Initially, D and G are assigned random weights. D is then trained with real and synthetic samples, while G is set as not trainable (fixed). Following the update of D's weights, it is marked as not trainable, and G is updated to minimize the loss from D, assuming that the G outputs are "real samples", Equation (2).

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{x \sim p_x(x)}[\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \end{aligned} \tag{2}$$

2.5.3. Conditional Generative Adversarial Networks

The conditional GAN (cGAN) is an extension of the previously presented GANs and was introduced in 2014 [8]. Unlike traditional GANs, conditional GANs have both the generative and discriminative networks conditioned on additional information, which learns the mapping from an observed image (x). This additional information helps to ensure that the images generated by the generator look as real as possible. In a cGAN, the generator function G maps a random noise vector (z) and an observed image (x) to the output vector (y) as follows: $G : \{x, z\} \rightarrow y$. The objective function of cGANs can be expressed as Equation (3).

$$\begin{aligned} \mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x \sim p_x(x)}[\log D(x, u)] \\ + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z, u)))] \end{aligned} \tag{3}$$

2.5.4. Pix2Pix-cGANs

Pix2Pix-cGANs were proposed to address image-to-image translation problems using the $L1$-distance [9], as shown in Equation (4).

$$\mathcal{L}_{L_1} = \mathbb{E}_{x \sim p_x(x)}[\|(x - G(u))\|_1] \tag{4}$$

This approach uses the input image to determine the output image. Pix2Pix-cGANs belong to the cGANs family, where G is trained with adversarial loss to generate a plausible image in the target domain. Additionally, using the specially designed G-D for conditional image classification, G is trained with additional loss to generate possible plausible translations. To update G, $L1$ loss is used, which measures the difference between the generated image and the predicted output.

2.5.5. Attention Mechanism for Pix2Pix-cGANs

In this study, an attention mechanism is added to both the generator and discriminator networks of the Pix2Pix model to enhance the quality of the image colorization process. This approach is inspired by our paper that compares attention-based Pix2Pix (Att Pix2Pix with Pix2Pix) [33], as shown in Figure 3.

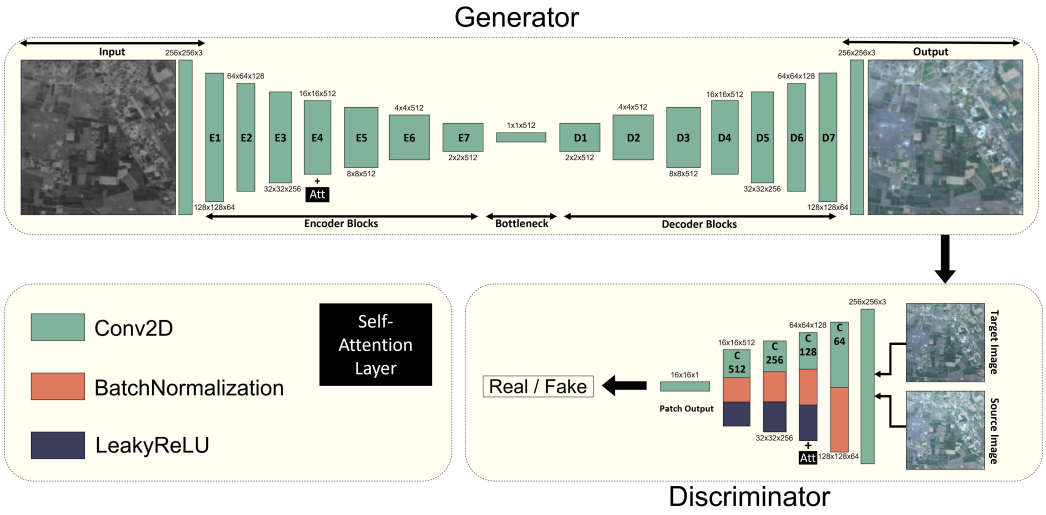

Figure 3. Architecture of improved Pix2Pix model with self-attention.

- Attention in the Generator: By integrating an attention mechanism into the generator, it allows selective focus on regions that require more detailed processing. Specifically, the attention mechanism helps the generator to detect areas with complex features, such as boundaries or intricate textures, where accurate colorization is more challenging. In addition, an attentive residual network was employed to guide the generator in preserving important low-level details while performing the colorization.
- Attention in the Discriminator: By incorporating an attention mechanism into the discriminator, it enhances the ability to focus on specific regions where the generated image might differ from the real image. This region-specific focus helps the discriminator to provide more targeted feedback to the generator, thereby improving the adversarial training process.

In this method, both a grayscale Landsat 5TM image and an RGB image serve as inputs, while an accurate RGB image represents the desired output.

2.5.6. BigColor

BigColor by [11] leverages pretrained generative adversarial networks (GANs) to synthesize vibrant and realistic image colorizations. The architecture employs a class-conditioned encoder–generator framework, where a spatial feature map extracted from a convolutional encoder is used as a latent representation for the generator. This setup enables the model to manage diverse and complex structures in grayscale images. The adversarial training alternates between optimizing the encoder–generator and the discriminator. The encoder–generator loss function is defined as Equation (5):

$$\mathcal{L}^G = \mathcal{L}^G_{\text{mse}} + \lambda_{\text{per}}\mathcal{L}^G_{\text{per}} + \lambda_{\text{adv}}\mathcal{L}^G_{\text{adv}} \tag{5}$$

where $\mathcal{L}^G_{\text{mse}}$ penalizes the mean squared error (MSE) between the synthesized image $\hat{x}_{\text{rgb}}$ and the ground truth x_{rgb}, and $\mathcal{L}^G_{\text{per}}$ is the perceptual loss computed using features from the 1st, 2nd, 6th, and 9th layers of the VGG16 network. The adversarial loss $\mathcal{L}^G_{\text{adv}}$ is defined as inEquation (6):

$$\mathcal{L}^G_{\text{adv}} = -D(\hat{x}_{\text{rgb}}, c) \tag{6}$$

where D represents the discriminator output and c is the class label. The balancing weights λ_{per} and λ_{adv} are empirically set to 0.2 and 0.03, respectively, as adopted in [11]. The discriminator is trained using the hinge loss, Equation (7):

$$\mathcal{L}_{\text{adv}}^{D} = -\min(0, -1 + D(x_{\text{rgb}}, c)) + \min(0, -1 - D(\hat{x}_{\text{rgb}}, c)) \tag{7}$$

This adversarial training framework ensures that the generator produces vibrant and perceptually realistic colors, while the discriminator evaluates the authenticity of the synthesized images.

2.5.7. ChromaGAN

ChromaGAN by [10] combines perceptual, semantic, and adversarial losses to produce realistic and semantically meaningful colorizations. The total loss function is defined as in Equation (8):

$$\mathcal{L}(G_\theta, D_w) = \mathcal{L}_e(G_{\theta_1}^1) + \lambda_g \mathcal{L}_g(G_{\theta_1}^1, D_w) + \lambda_s \mathcal{L}_s(G_{\theta_2}^2) \tag{8}$$

where $\mathcal{L}_e$ minimizes the error in chrominance values, $\mathcal{L}_g$ is the adversarial WGAN loss, and $\mathcal{L}_s$ aligns the semantic distributions. The hyperparameters λ_g and λ_s balance these terms. The color error loss ensures perceptual alignment between generated and ground truth chrominance values in the CIE Lab color space, Equation (9):

$$\mathcal{L}_e(G_{\theta_1}^1) = \mathbb{E}_{(L,a_r,b_r)\sim P_r}\left[\|G_{\theta_1}^1(L) - (a_r, b_r)\|_2^2\right] \tag{9}$$

The semantic loss aligns the predicted class distribution $G_{\theta_2}^2$ with a pretrained VGG-16 output y^v, Equation (10):

$$\mathcal{L}_s(G_{\theta_2}^2) = \mathbb{E}_{L\sim P_{r_g}}\left[\text{KL}\left(y_v \| G_{\theta_2}^2(L)\right)\right] \tag{10}$$

The WGAN loss, designed for stable adversarial training, is in Equation (11):

$$\begin{aligned}\mathcal{L}_g(G_{\theta_1}^1, D_w) = \mathbb{E}_{\tilde{I}\sim P_r}[\log D_w(\tilde{I})] - \mathbb{E}_{(a,b)\sim P_{G_{\theta_1}^1}}[\log D_w(L,a,b)] \\ -\mathbb{E}_{\hat{I}\sim P_{\hat{I}}}[(\|\nabla_{\hat{I}} D_w(\hat{I})\|_2 - 1)^2]\end{aligned} \tag{11}$$

The model optimizes the generator G_θ and discriminator D_w via a min–max objective, Equation (12):

$$\min_{G\theta} \max_{D_w \in \mathcal{D}} \mathcal{L}(G_\theta, D_w) \tag{12}$$

where $\mathcal{D}$ is the set of 1-Lipschitz functions. This framework ensures vivid, perceptually consistent, and semantically accurate results.

2.5.8. iColoriT

iColoriT by [12] model addresses point-interactive image colorization by leveraging a vision transformer architecture to propagate user hints globally across the image. Given a grayscale input $I_g \in \mathbb{R}^{H\times W\times 1}$ concatenated with sparse user hints $I_{\text{hint}} \in \mathbb{R}^{H\times W\times 3}$, the model predicts chrominance values in the CIE Lab color space. The input $X = I_g \oplus I_{\text{hint}}$ is reshaped into patches of size $P \times P$, where each patch is treated as a token. A sinusoidal positional encoding $E_{\text{pos}} \in \mathbb{R}^{N\times d}$ is added to incorporate spatial information, where $N = \frac{H\cdot W}{P^2}$ is the number of tokens. The transformer encoder processes the input tokens using multi-head self-attention (MSA) and feed-forward layers, as in Equations (13)–(15):

$$z_0 = X_p + E_{\text{pos}} \tag{13}$$

$$z_l' = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1} \tag{14}$$

$$z_l = \text{MLP}(\text{LN}(z'_l)) + z'_l \tag{15}$$

where $\text{LN}(\cdot)$ denotes layer normalization, and d is the hidden dimension. The attention mechanism computes similarities between all spatial locations, Equation (16):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{16}$$

where $Q, K, V \in \mathbb{R}^{N\times d}$ are query, key, and value matrices, and $B \in \mathbb{R}^{N\times N}$ is a relative positional bias. This enables efficient propagation of user hints to relevant regions at all layers. The output features from the transformer encoder $y_p \in \mathbb{R}^{N\times d}$ are reshaped into a feature map $y \in \mathbb{R}^{\frac{H}{P}\times\frac{W}{P}\times d}$. To upsample this feature map to the original resolution, iColoriT employs pixel shuffling Equation (17):

$$I_{\text{ab\_pred}} = \text{PS}(\text{LS}(y)) \tag{17}$$

where PS is pixel shuffling and LS is the local stabilizing layer. The local stabilizing layer restricts the receptive field of the last layer to mitigate artifacts caused by large upsampling ratios. The model minimizes the Huber loss in the CIE Lab color space, Equation (18):

$$\begin{aligned}\mathcal{L}_{\text{recon}} = &\frac{1}{2}(I_{\text{pred}} - I_{\text{GT}})^2 \mathbb{1}(|I_{\text{pred}} - I_{\text{GT}}| < 1) \\ &+ (|I_{\text{pred}} - I_{\text{GT}}| - \frac{1}{2}) \mathbb{1}(|I_{\text{pred}} - I_{\text{GT}}| \geq 1)\end{aligned} \tag{18}$$

2.5.9. Colorization Model Selection

Recent literature supports the selection of Att Pix2Pix-cGANs, BigColor, ChromaGAN, and iColoriT for satellite image colorization. These models represent state-of-the-art approaches that have been extensively compared or adopted in recent comparative studies. For example, ChromaGAN and BigColor have frequently been used as baseline models, highlighting the effectiveness of generating realistic colors in images [34]. The transformer-based model iColoriT uses a global receptive field to propagate user hints throughout the image; its selection in this study is motivated by its promising effectiveness despite limited prior research [35]. Furthermore, the Att Pix2Pix framework has proven particularly successful in the remote sensing context, significantly enhancing image translation quality through improved spatial attention mechanisms [33]. These studies collectively validate the models chosen as optimal for satellite imagery colorization tasks.

2.5.10. Setting Up the Colorization Models

Each model was implemented in Google Colab and trained for 100 epochs to transform grayscale Landsat 5TM images into accurate RGB outputs. The integration of attention mechanisms in Att Pix2Pix, alongside the unique approaches of BigColor, ChromaGAN, and iColoriT, aimed to enhance colorization quality, particularly for regions of interest, such as built-up areas.

All images were split into training (80%) and validation (20%) sets (1008 images for Vendée and 303 for Greater Agadir). All models were trained using the Adam optimizer, with a batch size of 16 and learning rate of 0.0002. Training on Google Colab's GPU took approximately 6 to 14 h per model. Each model's loss functions were retained as described: Att Pix2Pix used L1 and adversarial losses, BigColor combined MSE, perceptual, and adversarial losses, ChromaGAN balanced chrominance, semantic, and WGAN losses, and iColoriT minimized Huber loss in the CIE Lab color space.

2.5.11. Evaluate Colorization Models

The performance of the models for the colorization was assessed using two frequently used measures: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), following standard evaluation practices [36].

- PSNR: PSNR is a measure that compares the maximum possible power of an image's signal to the power of the noise that degrades it, expressed in decibels on a logarithmic scale, as in Equation (19). To find PSNR, the mean squared error (MSE) is first calculated, as in Equation (20), averaging the squared differences between the original and distorted image pixels using the image's dimensions of M rows and N columns.

$$\text{PSNR} = 10 \log_{10} \left(\frac{R^2}{\text{MSE}} \right) \tag{19}$$

$$\text{MSE} = \frac{1}{M \cdot N} \sum_{m,n} [I_1(m,n) - I_2(m,n)]^2 \tag{20}$$

- SSIM: SSIM assesses how similar two images are by analyzing their structural features, aligning with human visual perception. For images x (generated) and y (ground truth), it uses their mean values (μ_x, μ_y), standard deviations (σ_x, σ_y), and covariance (σ_{xy}), as shown in Equation (21):

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{21}$$

  Here, C_1 and C_2 are small constants to prevent division by zero, typically $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$, where L is the maximum pixel value.

2.6. Training Models for Segmentation

2.6.1. Training Data for Segmentation

To generate suitable input–output pairs for training, the data underwent a series of specific preprocessing steps for the model:

- Patching: The colorized Landsat images taken from the previous best colorization model and built-up masks were initially padded to ensure that their dimensions were divisible by the specified patch size. The images were divided into patches of size 256 × 256 pixels.
- Data Augmentation: Data augmentation techniques were used, including rotations, translations, horizontal flips, shear transformations, and zooming.

2.6.2. SegFormer

SegFormer is a semantic segmentation architecture introduced by [16], distinguished by its two primary components: a hierarchical transformer encoder and a lightweight All-MLP decoder. This architecture is designed to efficiently handle semantic segmentation tasks by leveraging multi-scale feature extraction and streamlined decoding processes, offering a balance between performance and computational efficiency.

The encoder in SegFormer is built around a series of mix transformer encoders, denoted as MiT-B0 through MiT-B5, which vary in size and capacity. Unlike the vision transformer (ViT), which produces uniform-scale features, SegFormer's encoder adopts a hierarchical structure that generates multi-level multi-scale feature representations from the input image. This approach enables the extraction of high-resolution coarse features as well as low-resolution fine-grained features, both of which are critical for achieving high-quality semantic segmentation. To address the computational complexity of traditional self-attention mechanisms, especially when applied to high-resolution feature maps, Seg-

Former incorporates an efficient self-attention mechanism paired with a sequence reduction strategy. This reduces the computational burden while preserving the model's ability to capture long-range dependencies effectively.

The efficient self-attention mechanism is mathematically defined as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{\text{head}}}}\right)V \tag{22}$$

where Q, K, and V represent the query, key, and value matrices, and d_{head} is the dimension of each attention head. To further optimize this, a sequence reduction process is introduced, reducing the length of the key sequence by a reduction ratio R. This is expressed in two steps:

$$\hat{K} = \text{Reshape}\left(\frac{N}{R}, C \cdot R\right)(K) \tag{23}$$

$$K' = \text{Linear}(C \cdot R, C)(\hat{K}) \tag{24}$$

Here, N is the original sequence length, C is the channel dimension, and K' is the reduced key sequence used in the attention computation. This reduction significantly lowers the memory and computational requirements, making SegFormer scalable for high-resolution inputs.

Another key feature of the encoder is its overlapped patch merging technique, which replaces the non-overlapping patch splitting of traditional transformers. This process is parameterized by K (patch size), S (stride), and P (padding size), allowing the model to maintain local spatial continuity, an essential aspect for segmentation tasks that rely on detailed spatial information. Unlike most transformer-based models, SegFormer eliminates explicit positional encoding. Instead, it implicitly conveys positional information through a 3×3 convolution within the Mix-FFN (feed-forward network) module, formulated as

$$x_{\text{out}} = \text{MLP}(\text{GELU}(\text{Conv}_{3\times3}(\text{MLP}(x_{\text{in}})))) + x_{\text{in}} \tag{25}$$

where x_{in} is the input feature, $\text{Conv}_{3\times3}$ is a 3×3 convolution, and GELU is the Gaussian Error Linear Unit activation function. This positional-encoding-free design simplifies the architecture while maintaining robust feature representation.

The decoder of SegFormer is notably lightweight, relying entirely on Multi-Layer Perceptron (MLP) layers rather than the complex convolutional or attention-based components found in traditional semantic segmentation models. This efficiency is made possible by the hierarchical transformer encoder, which provides a larger effective receptive field (ERF) compared to conventional CNN-based encoders. The decoding process consists of four key steps: (1) multi-level features extracted from the MiT encoder (at different scales) are processed by an MLP layer to standardize their channel dimensions; (2) these features are upsampled to a uniform resolution, specifically a quarter of the original input resolution, and concatenated; (3) a second MLP layer fuses the concatenated features into a cohesive representation; (4) finally, an additional MLP layer generates the segmentation mask, predicting class labels at a quarter of the original resolution.

2.6.3. U-Net++

U-Net++ by [15] is a deeply supervised encoder–decoder architecture designed for medical image segmentation. It extends the original U-Net by introducing nested and dense skip pathways that connect the encoder and decoder subnetworks. These redesigned skip pathways aim to reduce the semantic gap between the feature maps of the encoder and decoder, simplifying the optimization task for the model.

The architecture is defined by a series of nested dense convolutional blocks along the skip pathways. Formally, let $x^{i,j}$ denote the output of node $X^{i,j}$, where i indexes the downsampling layer along the encoder and j indexes the convolution layer of the dense block along the skip pathway. The stack of feature maps represented by $x^{i,j}$ is computed as in Equation (26):

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), & j = 0 \\ H([x^{i,k}]_{k=0}^{j-1}, U(x^{i+1,j-1})), & j > 0 \end{cases} \tag{26}$$

where $H(x^{i-1,j})$ is a convolution operation followed by an activation function, $U(x^{i+1,j-1})$ denotes an upsampling layer, and $[x^{i,k}]$ represents the concatenation layer. Nodes at level $j = 0$ receive only one input from the previous layer of the encoder, while nodes at higher levels ($j > 0$) aggregate outputs from all prior nodes in the same skip pathway and the upsampled output from the lower skip pathway.

The nested skip pathways enable gradual enrichment of high-resolution feature maps from the encoder before their fusion with semantically rich feature maps from the decoder. This design ensures that the feature maps from the encoder and decoder are semantically similar, making the optimization problem easier for the model.

In addition to the redesigned skip pathways, U-Net++ incorporates deep supervision, which allows the model to operate in two modes:

- Accurate Mode: Outputs from all segmentation branches are averaged to produce the final segmentation map.
- Fast Mode: The final segmentation map is selected from only one of the segmentation branches, enabling model pruning and reducing inference time.

The loss function for deep supervision combines binary cross-entropy and the Dice coefficient, defined as in Equation (27):

$$\mathcal{L}(Y, \hat{Y}) = -\frac{1}{N} \sum_{b=1}^{N} \left(\frac{1}{2} \cdot Y_b \cdot \log \hat{Y}_b + \frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b} \right) \tag{27}$$

where $\hat{Y}_b$ and Y_b denote the flattened predicted probabilities and ground truth labels for the b-th image, respectively, and N is the batch size.

2.6.4. DeepLabv3+

DeepLabv3+ by [14] is an advanced encoder–decoder architecture designed for semantic image segmentation. It extends DeepLabv3 by incorporating a simple yet effective decoder module to refine object boundaries while leveraging the rich contextual information encoded by the encoder. The architecture combines spatial pyramid pooling and encoder–decoder structures, enabling capture of multi-scale contextual information and recovery of sharp object boundaries.

The encoder employs atrous convolution to extract features at multiple scales, allowing explicit control of feature map resolution. The Atrous Spatial Pyramid Pooling (ASPP) module probes incoming features with filters at multiple rates and effective fields of view, capturing rich contextual information. The decoder refines segmentation results by gradually recovering spatial information and sharpening object boundaries.

The atrous convolution operation is defined as in Equation (28):

$$y[i] = \sum_{k} x[i + r \cdot k] w[k] \tag{28}$$

where x is the input feature map, w is the convolution filter, r is the atrous rate, and y is the output feature map. This generalizes standard convolution, enabling adjustment of the filter's field of view to capture multi-scale information.

To reduce computational complexity, DeepLabv3+ adopts depthwise separable convolution, factorizing standard convolution into a depthwise convolution followed by a pointwise convolution.

2.6.5. FPN

Feature Pyramid Networks (FPNs) by [13] are a generic framework addressing multi-scale object detection in deep convolutional neural networks. FPN leverages the inherent multi-scale pyramidal hierarchy of ConvNets to construct feature pyramids with marginal extra cost. The architecture combines low-resolution, semantically strong features with high-resolution, semantically weak features through a top-down pathway and lateral connections.

The FPN architecture consists of

1. Bottom-up pathway: Feed-forward computation of the backbone ConvNet, computing a feature hierarchy at several scales with a scaling step of 2.
2. Top-down pathway and lateral connections: Higher-resolution features are generated by upsampling coarser, semantically stronger feature maps from higher pyramid levels, enhanced with bottom-up pathway features via lateral connections, as in Equation (29):

$$P_l = \text{Conv}_{3\times3}\Big(\text{Resize}(P_{l+1}) + \text{Conv}_{1\times1}(C_l)\Big) \tag{29}$$

3. Prediction heads: A shared prediction head is attached to each pyramid level.

For an RoI with width w and height h, the level P_k is determined by Equation (30):

$$k = \lfloor k_0 + \log_2(\sqrt{wh}/224) \rfloor \tag{30}$$

where k_0 is the target level for an RoI with $w \times h = 224^2$.

2.6.6. Segmentation Model Selection

The choice of SegFormer, U-Net++, DeepLabv3+, and FPN for satellite image segmentation is strongly supported by recent comparative research. SegFormer, a transformer-based architecture, has achieved the highest segmentation accuracy in recent remote sensing evaluations, particularly excelling at capturing fine-scale features [37]. DeepLabv3+ consistently appears as a strong benchmark, excelling in segmentation precision in complex land cover classification tasks [38]. U-Net++, recognized for its enhanced multi-scale segmentation capabilities, remains widely used in remote sensing studies due to its consistently high performance on diverse segmentation challenges [39]. Finally, FPN-based architectures have been extensively validated in remote sensing scenarios, effectively improving multi-scale feature extraction and segmentation accuracy [40]. These recent findings reinforce the suitability of the selected segmentation models for the task.

2.6.7. Setting Up Segmentation Models

Each model was implemented using the Segmentation Models PyTorch(2.6) (SMP) library and trained for 100 epochs to segment grayscale and colorized Landsat 5TM images into built-up-area outputs. The hierarchical transformer-based approach of SegFormer, the nested skip connections of U-Net++, the atrous convolutions of DeepLabv3+, and the feature pyramid structure of FPN were leveraged to improve segmentation accuracy for complex regions, such as built-up areas.

All images were divided into training (80%) and validation (20%) sets, comprising 1008 images for Vendée and 303 for Greater Agadir. The models were trained using the Adam optimizer on Google Colab's GPU, with training times ranging from approximately 3 to 6 h per model, depending on the architecture complexity and encoder size. Each model's loss functions were retained as per their standard formulations: SegFormer used a combination of cross-entropy and Dice losses, U-Net++ employed a weighted cross-entropy loss with deep supervision, DeepLabv3+ utilized cross-entropy with an auxiliary loss, and FPN minimized a standard cross-entropy loss. All models were pretrained on ResNet101. However, Segformer was pretrained on MiT-B5.

2.6.8. Evaluating Segmentation Models

The performance of the segmentation models was assessed using six widely adopted metrics: accuracy, precision, recall, F1 score, Intersection over Union (IoU), and Mean Intersection over Union (mIoU), following standard evaluation practices in semantic segmentation [41].

- Accuracy: Accuracy measures the proportion of correctly classified pixels across all classes in the image. It is defined as the ratio of true predictions (both positive and negative) to the total number of pixels, expressed in Equation (31), where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{31}$$

- Precision: Precision quantifies the accuracy of positive predictions, representing the ratio of correctly predicted positive pixels to the total predicted positive pixels. It is calculated as shown in Equation (32).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{32}$$

- Recall: Recall, or sensitivity, measures the ability of the model to identify all relevant positive pixels. It is the ratio of correctly predicted positive pixels to the total actual positive pixels, defined in Equation (33).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{33}$$

- F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance. It is particularly useful when class distribution is imbalanced, as shown in Equation (34).

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{34}$$

- IoU: Intersection over Union (IoU), also known as the Jaccard Index, calculates the overlap between the predicted segmentation mask and the ground truth by dividing the area of their intersection by the area of union. Here, A represents the predicted pixel set, and B represents the ground truth pixel set for a given class. It is expressed in Equation (35):

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{35}$$

- mIoU: Mean Intersection over Union (mIoU) averages the IoU scores across all classes in the dataset, providing a comprehensive evaluation of segmentation performance.

For N classes, A_i and B_i denote the predicted and ground truth pixel sets for the i-th class, respectively. It is defined in Equation (36):

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \tag{36}$$

3. Results

3.1. Evaluation of Colorization Methods

This study evaluated the performance of four distinct colorization methods on historical grayscale satellite images from two regions: Greater Agadir and Vendée, using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) as evaluation metrics. The results, illustrated in Figure 4, demonstrate that Att Pix2Pix consistently outperformed the other methods across all the images, as validated in our research [33].

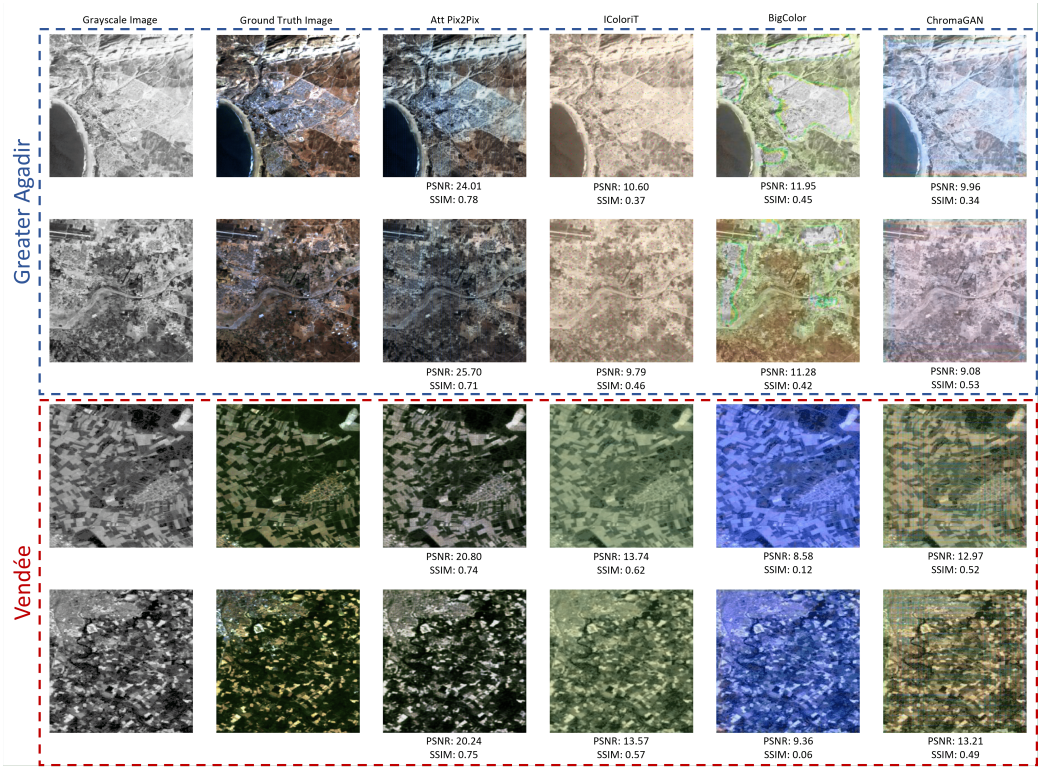

Figure 4. Comparison of different colorization methods on satellite images from Greater Agadir and Vendée regions.

For Greater Agadir, the first image, Att Pix2Pix, achieved a PSNR of 24.01 and SSIM of 0.78, excelling in noise reduction and structural similarity, with realistic colors for urban, water, vegetation, and bareland features. iColoriT had a PSNR of 10.60 and SSIM of 0.37, struggling with urban areas due to muted and inconsistent colors. BigColor showed a PSNR of 11.95 and SSIM of 0.45, performing well in barelands but with color confusion in water, built-up areas, and vegetation. ChromaGAN had a PSNR of 9.96 and SSIM of 0.34, the lowest quantitatively, but visually closer to Att Pix2Pix. The second image, Att Pix2Pix, maintained high performance with a PSNR of 25.70 and SSIM of 0.71, realistic across diverse features. iColoriT had a PSNR of 9.79 and SSIM of 0.46, again struggling in urban areas. BigColor showed a PSNR of 11.28 and SSIM of 0.42, similar trends. ChromaGAN improved slightly with a PSNR of 9.08 and SSIM of 0.53, resembling Att Pix2Pix in urban and coastal areas.

In the Vendée region, similar trends were observed. For the first image, Att Pix2Pix achieved a PSNR of 20.80 and SSIM of 0.74, robust across water, vegetation, barelands, and built-up areas. iColoriT had a PSNR of 13.74 and SSIM of 0.62, reasonable but less vibrant in urban areas. BigColor showed a PSNR of 8.58 and SSIM of 0.12, poor performance with significant color confusion. ChromaGAN improved, with a PSNR of 12.97 and SSIM of 0.52, visually close to Att Pix2Pix in coastal and vegetated regions. The second image maintained the Att Pix2Pix values, led by a PSNR of 20.24 and SSIM of 0.75, precise colorization. iColoriT had a PSNR of 13.57 and SSIM of 0.57, facing urban challenges. BigColor showed a PSNR of 9.36 and SSIM of 0.06, very low, with poor structural preservation. ChromaGAN maintained a PSNR of 13.21 and SSIM of 0.49, similar trends.

To provide a comprehensive overview, the average PSNR and SSIM values for each method were calculated separately for Greater Agadir and Vendée, based on images of validation per region, as shown in Table 3 and Figure 5:

Table 3. Average PSNR and SSIM for colorization methods in Greater Agadir and Vendée.

Model	Greater Agadir		Vendée	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Att Pix2Pix	27.18	0.95	27.72	0.96
iColoriT	23.87	0.88	25.33	0.84
BigColor	11.92	0.56	12.11	0.58
ChromaGAN	23.44	0.85	22.64	0.79

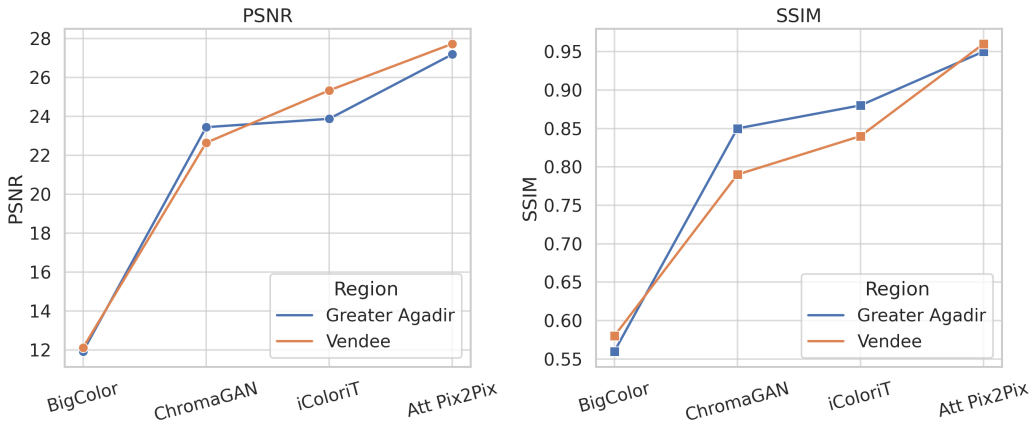

Figure 5. PSNR and SSIM values for each colorization model across two study regions (Greater Agadir and Vendée), ordered by average overall performance.

The consistent superiority of Att Pix2Pix across both regions is evident from its highest PSNR and SSIM values: 27.18 and 0.95 in Greater Agadir and 27.72 and 0.96 in Vendée. This performance likely stems from its attention mechanism, which effectively captures and reproduces accurate details and structures. The regional differences in model performance suggest that the image characteristics influence the effectiveness of the colorization methods. iColoriT shows a higher PSNR in Vendée (25.33) compared to Greater Agadir (23.87), although its SSIM is slightly lower (0.84 vs 0.88). BigColor performs the lowest in both regions. ChromaGAN, however, has a lower PSNR and SSIM in Vendée compared to Greater Agadir. These findings support our selection of Att Pix2Pix for colorization as it consistently delivers high-quality results across both regions, crucial for enhancing segmentation accuracy, as validated in our research [33]. Additionally, the user observed that ChromaGAN's colors visually resemble those of Att Pix2Pix despite lower metrics, indicating a potential discrepancy between the numerical and perceptual quality assessments.

3.2. Evaluation of Segmentation Models

The evaluation compares models trained on grayscale versus colorized datasets from two distinct regions: Greater Agadir and Vendée, utilizing a comprehensive set of metrics including accuracy, precision, recall, F1 score, Intersection over Union (IoU), and mean Intersection over Union (mIoU) (Figures 6 and 7).

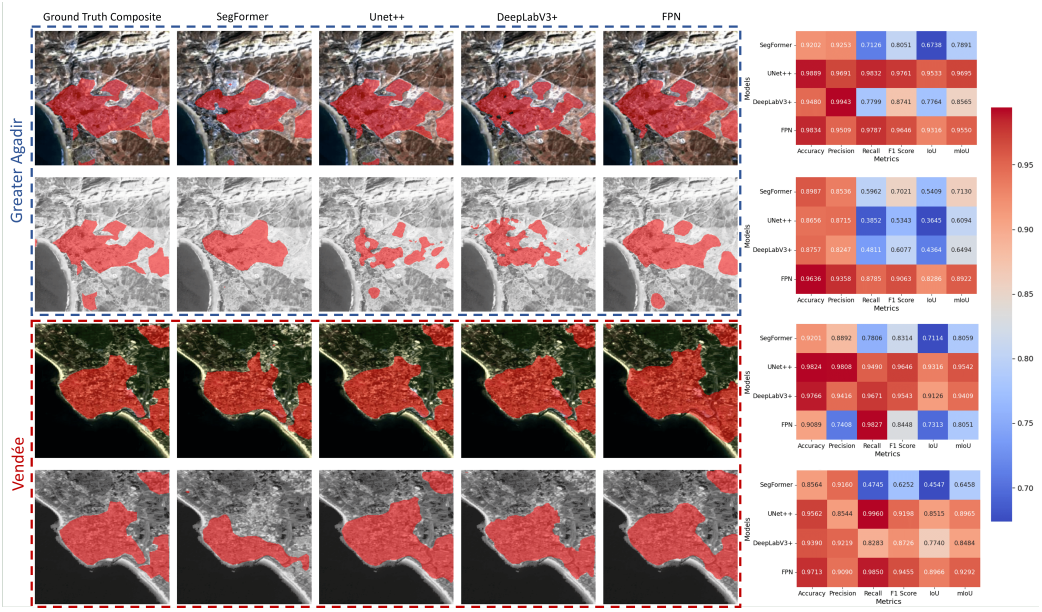

Figure 6. Comparison of segmentation results for different models on satellite images from Greater Agadir and Vendée regions.

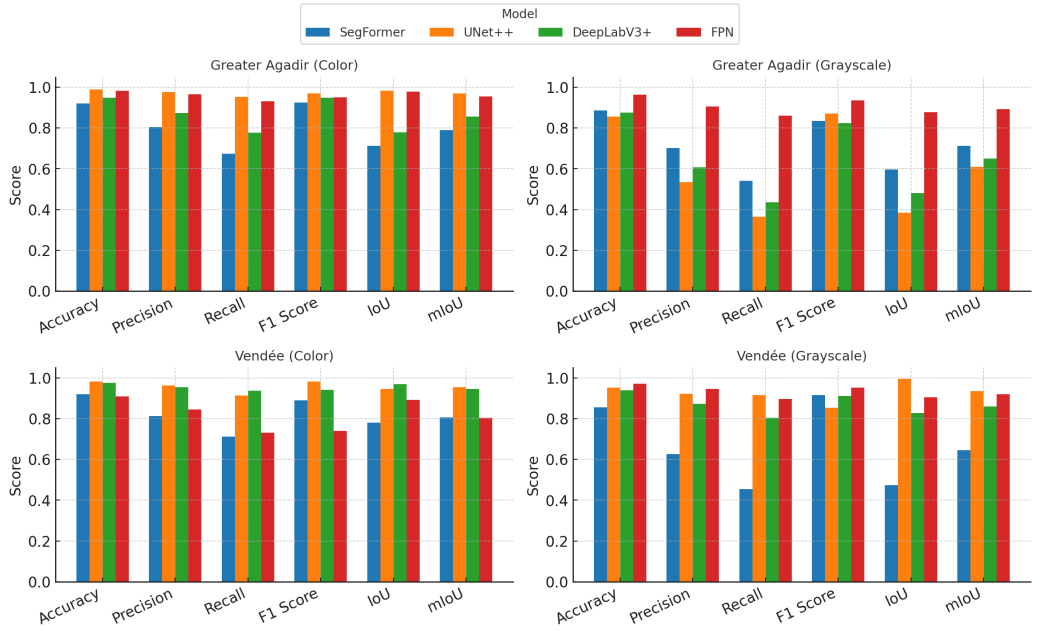

Figure 7. Segmentation scores for four models across colorized and grayscale satellite images from Greater Agadir and Vendée. Metrics include accuracy, precision, recall, F1 score, IoU, and mIoU. Colors represent different models.

For Greater Agadir, the results indicate a marked improvement in segmentation performance when the models are trained on colorized images compared to grayscale counterparts. Specifically, UNet++ achieved an accuracy of 86.56%, F1 score of 53.43%, and mIoU of 60.94% with grayscale training, but colorized training significantly boosted these to 98.89%, 97.61%, and 99.55%, respectively, demonstrating exceptional performance

gains, particularly in detecting built-up areas. FPN recorded an accuracy of 96.36%, F1 score of 90.63%, and mIoU of 82.22% with grayscale training, further elevated to 98.34%, 96.46%, and 96.50% with colorized training, showcasing superior segmentation capabilities. The images provided clearly illustrate that the models struggle to segment built-up areas in grayscale images, with visible gaps and errors, whereas colorized images enable better detection, particularly for urban features, compared to the ground truth images.

Comparable trends emerged in the Vendée region, where the colorized-trained models consistently outperformed their grayscale-trained counterparts across all the metrics. UNet++ attained an accuracy of 95.62%, F1 score of 91.98%, and mIoU of 89.56% with grayscale training, elevated to 98.24%, 96.46%, and 95.42% with colorized training, demonstrating significant performance enhancement in detecting built-up areas. FPN achieved an accuracy of 97.13%, F1 score of 94.55%, and mIoU of 92.92% with grayscale training, but colorized training resulted in an accuracy of 90.89%, F1 score of 84.48%, and mIoU of 80.51%, possibly due to regional image characteristics affecting generalization. The figures for Vendée also reveal that the grayscale-trained models struggled with segmentation, particularly in the river area, where it is clear that the model generalized it with built-up areas, while the colorized-trained models, aided by colorization, showed improved detection of built-up areas.

4. Discussion

The application of Att Pix2Pix for colorization and RGB-trained UNet++ for segmentation to historical images of Agadir and Les Sables d'Olonne confirms the effectiveness of the process, as visualized in Figure 8. The colorized images show high fidelity, and the segmented outputs demonstrate improved accuracy for built-up areas, highlighting the critical role of colorization in enhancing segmentation outcomes.

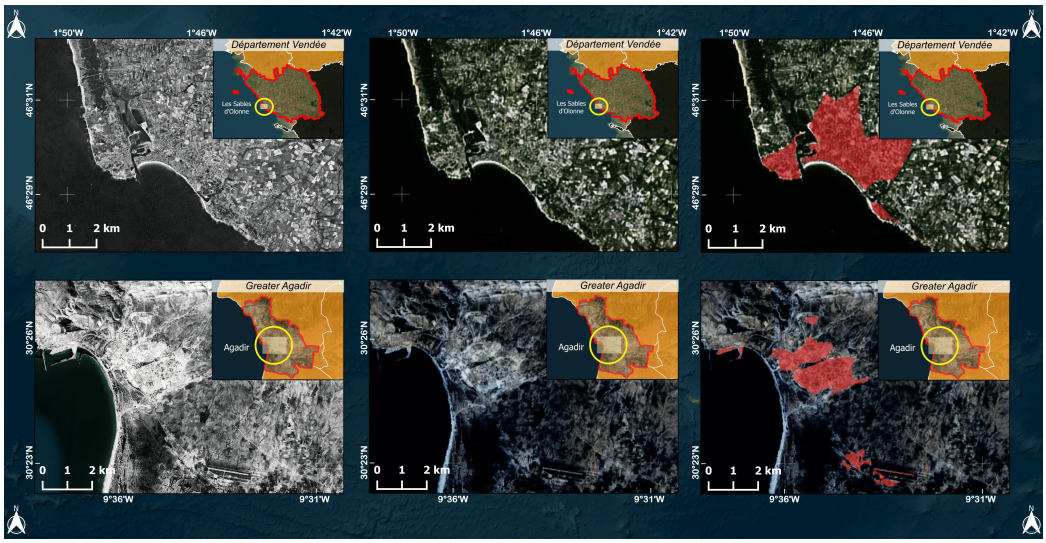

Figure 8. Colorization and segmentation of Agadir in 1967 and Les Sables d'Olonne in 1975.

In the initial evaluations, the segmentation of grayscale images using grayscale-trained models was compared to RGB-trained models on colorized images, finding that RGB-trained models, particularly UNet++, consistently outperformed, with higher accuracy. This comparison underscores that segmenting colorized images is superior to directly segmenting grayscale images as color information provides additional spectral cues for distinguishing land cover types, especially in complex urban areas. This is why this study was conducted: to prove that colorizing first with Att Pix2Pix and then segmenting with

RGB-trained UNet++ is better than segmenting grayscale images directly, addressing the limitations of grayscale data in historical imagery analysis.

For the test cases of Agadir and Sables d'Olonne, the grayscale archive images were colorized using Att Pix2Pix and then segmented with the RGB-trained UNet++, achieving results consistent with the initial findings. The figure illustrates the original grayscale image, the colorized image, and the segmented output, visually demonstrating the improvement in segmentation accuracy when using colorized images for built-up areas compared to the known struggles of grayscale-trained models, as established in the earlier evaluations. This approach validates that integrating colorization enhances segmentation, offering a practical solution for historical LULC analysis.

While the current study demonstrates the effectiveness of colorizing historical grayscale images using Att Pix2Pix followed by segmentation with RGB-trained UNet++ for built-up areas in Agadir and Les Sables d'Olonne, several limitations must be acknowledged to guide future research. First, the use of higher-resolution imagery could potentially enhance segmentation accuracy, as evidenced by studies showing improved land use classification with a higher spatial resolution [42]. However, processing high-resolution images demands substantial computational resources and time, which may not be practical for large-scale applications without access to advanced computing infrastructure, a concern particularly relevant for extensive areas. Second, the model's adaptability to more complex scenarios, such as areas with diverse land cover types or varying urban densities, requires further evaluation. The current focus on built-up areas might limit the model's performance in regions with mixed land uses or complex urban patterns, and training on a wider variety of land use types could improve the strength and validity, potentially leveraging transfer learning from models pretrained on large multiclass land cover datasets like BigEarthNet [43]. Additionally, the dependency on extensive labeled data for training supervised models like UNet++ poses a challenge, particularly for historical imagery where ground truth is hard to find. Future research could explore semisupervised or unsupervised learning techniques to reduce the need for labeled data. Furthermore, the accuracy of the colorization step is crucial as errors in Att Pix2Pix can impact the segmentation stage, necessitating methods to quantify and mitigate these errors. Historical imagery often suffers from degradation noise, blurriness, or low contrast, which may compromise model performance, suggesting a need for strong image restoration techniques. Lastly, the computational costs associated with training and deploying these deep learning models are significant, and optimizing model architectures for efficiency or employing model compression techniques could make the approach more accessible for larger applications. These considerations, including exploring higher resolutions and diverse land use training, are integral to the ongoing development of methods for historical LULC analysis, ensuring that future studies can build upon the current findings to achieve more accurate and efficient results as part of a larger scope.

5. Conclusions

The integrated approach of colorizing historical grayscale satellite images using Att Pix2Pix and then segmenting with RGB-trained UNet++ provides a robust solution for accurate land use and land cover (LULC) analysis, particularly for built-up-area detection. This methodology significantly enhances segmentation accuracy by leveraging color information, as demonstrated by the superior performance of RGB-trained models compared to grayscale-trained ones, validating the hypothesis that colorization is essential for effective historical imagery analysis. The successful application to Agadir in 1967 and Sables d'Olonne in 1975, based on high-performance metrics in Greater Agadir and

Vendée, confirms the beneficial implications for urban planning, environmental conservation, and disaster preparation.

Quantitative evaluations further support these findings. For colorization, Att Pix2Pix achieved an average PSNR of 27.18 and SSIM of 0.95 in Greater Agadir, and 27.72 and 0.96 in Vendée, consistently outperforming other methods such as iColoriT, BigColor, and ChromaGAN. In segmentation, RGB-trained UNet++ on colorized images reached an accuracy of 98.89%, F1 score of 97.61%, and mIoU of 99.55% in Greater Agadir, significantly surpassing the grayscale-trained models, which achieved an accuracy of 86.56%, F1 score of 53.43%, and mIoU of 60.94%. Similar improvements were observed in Vendée, with accuracy rising from 95.62% to 98.24% and mIoU from 89.56% to 95.42% when using colorized images. These results underscore the practical benefits of integrating colorization into historical LULC analysis, enhancing the ability to detect built-up areas critical for applications in urban planning and environmental monitoring.

Future research should focus on expanding this methodology to diverse regions, refining evaluation metrics, and addressing region-specific performance variations. Exploring higher-resolution imagery, adapting the approach to complex land cover types, and optimizing computational efficiency could further enhance its applicability and robustness, ensuring broader utility in historical satellite imagery analysis.

Author Contributions: Conceptualization, M.R.S., M.M. (Mohamed Maanan), M.M. (Mehdi Maanan), and H.R.; methodology, M.R.S., M.M. (Mohamed Maanan), and H.R.; software, M.R.S.; validation, M.R.S. and S.L.; formal analysis, M.R.S. and S.L.; investigation, M.R.S.; resources, M.R.S.; data curation, M.R.S. and S.L.; writing—original draft preparation, M.R.S.; writing—review and editing, M.R.S., M.M. (Mohamed Maanan), M.M. (Mehdi Maanan), and H.R.; visualization, M.R.S.; supervision, M.M. (Mohamed Maanan), M.M. (Mehdi Maanan), and H.R.; project administration, M.R.S. and H.R.; funding acquisition, M.M. (Mohamed Maanan) and M.M. (Mehdi Maanan). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this article are available upon request from the corresponding author.

Acknowledgments: Mohamed Rabii Simou acknowledges the Centre National pour la Recherche Scientifique et Technique(CNRST), Kingdom of Morocco, for the "Ph.D. ASsociate Scholarship-PASS".

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Att Pix2Pix	Attention-based Pix2Pix
cGAN	Conditional Generative Adversarial Network
FPN	Feature Pyramid Network
GAN	Generative Adversarial Network
GEE	Google Earth Engine
HCP	Haut Commissariat au Plan
INSEE	Institut National de la Statistique et des Études Économiques
IoU	Intersection over Union
LULC	Land Use and Land Cover
mIoU	Mean Intersection over Union
MSE	Mean Squared Error
PSNR	Peak Signal-to-Noise Ratio
RGB	Red, Green, Blue

RGPH	Recensement Général de la Population et de l'Habitat
SSIM	Structural Similarity Index
TOA	Top of Atmosphere
UNet++	Enhanced U-Net
USGS	United States Geological Survey

References

1. Jensen, J.R. *Remote Sensing of the Environment: An Earth Resource Perspective 2/e*; Pearson Education India: Tamil Nadu, India, 2009.
2. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
3. Library of Congress. History of Remote Sensing. 2024. Available online: https://guides.loc.gov/geospatial/computer-cartography-archive/history-of-remote-sensing (accessed on 13 February 2025).
4. Williams, D.L.; Goward, S.; Arvidson, T. Landsat. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1171–1178. [CrossRef]
5. National Reconnaissance Office. The CORONA Story. 1995. Available online: https://www.nro.gov/Portals/65/documents/history/csnr/corona/The%20CORONA%20Story.pdf (accessed on 10 February 2025).
6. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
7. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
8. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
9. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
10. Vitoria, P.; Raad, L.; Ballester, C. Chromagan: Adversarial picture colorization with semantic class distribution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2445–2454.
11. Kim, G.; Kang, K.; Kim, S.; Lee, H.; Kim, S.; Kim, J.; Baek, S.H.; Cho, S. Bigcolor: Colorization using a generative color prior for natural images. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 350–366.
12. Yun, J.; Lee, S.; Park, M.; Choo, J. iColoriT: Towards propagating local hints to the right region in interactive colorization by leveraging vision transformer. In Proceedings of the IEEE/CVF winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 1787–1796.
13. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
14. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
15. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
16. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
17. Lee, C. Artificial Colorization of Grayscale Satellite Imagery via GANs: Part 1. 2017. Available online: https://medium.com/the-downlinq/artificial-colorization-of-grayscale-satellite-imagery-via-gans-part-1-79c8d137e97b (accessed on 20 January 2025).
18. Gravey, M.; Rasera, L.G.; Mariethoz, G. Analogue-based colorization of remote sensing images using textural information. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 242–254. [CrossRef]
19. Liu, H.; Fu, Z.; Han, J.; Shao, L.; Liu, H. Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks. *J. Vis. Commun. Image Represent.* **2018**, *53*, 20–30. [CrossRef]
20. Agapiou, A. Land cover mapping from colorized CORONA archived greyscale satellite data and feature extraction classification. *Land* **2021**, *10*, 771. [CrossRef]
21. Farella, E.M.; Malek, S.; Remondino, F. Colorizing the past: Deep learning for the automatic colorization of historical aerial images. *J. Imaging* **2022**, *8*, 269. [CrossRef] [PubMed]
22. Shamsaliei, S.; Gundersen, O.E.; Alfredsen, K.T.; Halleraker, J.H.; Foldvik, A. Highlighting Challenges of State-of-the-Art Semantic Segmentation with HAIR-A Dataset of Historical Aerial Images. *J. Data Centric Mach. Learn. Res.* **2024**, 8, 1−31.
23. Anwar, S.; Tahir, M.; Li, C.; Mian, A.; Khan, F.S.; Muzaffar, A.W. Image colorization: A survey and dataset. *Inf. Fusion* **2025**, *114*, 102720. [CrossRef]

24. Haut Commissariat au Plan (HCP). Résultats du Recensement Général de la Population et de l'Habitation 2024 (RGPH 2024). 2024. Available online: https://resultats2024.rgphapps.ma/ (accessed on 22 February 2025).
25. Commune d'Agadir. Plan d'Action Communal d'Agadir 2022–2027. Technical Report, Commune d'Agadir. 2022. Available online: https://agadir2027.ma/wp-content/uploads/2023/04/Version-finale-du-Plan-dAction-Communal-2022-2027-1.pdf (accessed on 22 February 2025).
26. United States Geological Survey (USGS). Impact of the 1960 Agadir Earthquake. 2023. Available online: https://earthquake.usgs.gov/earthquakes/eventpage/iscgem878424/impact, (accessed on 22 February 2025).
27. Institut National de la Statistique et des Études Économiques (INSEE). Population Data for Les Sables-d-Olonne 2021. 2025. Available online: https://www.insee.fr/fr/statistiques/2011101?geo=COM-85194 (accessed on 23 February 2025).
28. Institut National de la Statistique et des Études Économiques (INSEE). Population Data for Vendée 2021. 2025. Available online: https://www.insee.fr/fr/statistiques/2011101?geo=DEP-85 (accessed on 23 February 2025).
29. Météo France. Fiche du Poste 85060002—Château-d'Olonne. 2024. Available online: https://donneespubliques.meteofrance.fr/metadonnees_publiques/fiches/fiche_85060002.pdf (accessed on 24 February 2025).
30. Vendée Globe. Vendée Globe Official Website. 2025. Available online: https://www.vendeeglobe.org/ (accessed on 25 February 2025).
31. Kanan, C.; Cottrell, G.W. Color-to-grayscale: Does the method matter in image recognition? *PLoS ONE* **2012**, *7*, e29740. [CrossRef] [PubMed]
32. Bourke, P. Histogram Matching. 2011. Available online: https://paulbourke.net/miscellaneous/equalisation/ (accessed on 13 December 2024).
33. Simou, M.R.; Maanan, M.; Loulad, S.; Benayad, M.; Maanan, M.; Rhinane, H. An Improved Pix2Pix Approach for Colorizing Historical Grayscale Satellite Imagery. 2025, *submitted*.
34. Shafiq, H.; Lee, B. Transforming color: A novel image colorization method. *Electronics* **2024**, *13*, 2511. [CrossRef]
35. Tran, D.T.; Nguyen, N.D.H.; Pham, T.T.; Tran, P.N.; Vu, T.D.T.; Nguyen, C.T.; Dang-Ngoc, H.; Dang, D.N.M. SwinTExCo: Exemplar-based video colorization using Swin Transformer. *Expert Syst. Appl.* **2025**, *260*, 125437. [CrossRef]
36. Bovik, A.C. *The Essential Guide to Image Processing*; Academic Press: Cambridge, MA, USA, 2009.
37. Lin, X.; Cheng, Y.; Chen, G.; Chen, W.; Chen, R.; Gao, D.; Zhang, Y.; Wu, Y. Semantic segmentation of China's coastal wetlands based on Sentinel-2 and Segformer. *Remote Sens.* **2023**, *15*, 3714. [CrossRef]
38. Feng, X.; Wei, C.; Xue, X.; Zhang, Q.; Liu, X. RST-DeepLabv3+: Multi-Scale Attention for Tailings Pond Identification with DeepLab. *Remote Sens.* **2025**, *17*, 411. [CrossRef]
39. Şengül, G.S.; Sertel, E. Automatic Building Extraction From VHR Remote Sensing Images Using Geoai Methods. In Proceedings of the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; IEEE: New York, NY, USA, 2024; pp. 8109–8112.
40. de Carvalho, O.L.F.; de Carvalho Júnior, O.A.; Silva, C.R.e.; de Albuquerque, A.O.; Santana, N.C.; Borges, D.L.; Gomes, R.A.T.; Guimarães, R.F. Panoptic segmentation meets remote sensing. *Remote Sens.* **2022**, *14*, 965. [CrossRef]
41. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
42. Fisher, J.R.; Acosta, E.A.; Dennedy-Frank, P.J.; Kroeger, T.; Boucher, T.M. Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sens. Ecol. Conserv.* **2018**, *4*, 137–149. [CrossRef]
43. Sumbul, G.; Charfuelan, M.; Demir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: New York, NY, USA, 2019; pp. 5901–5904.