Transforming National Air Photo Archives into Analysis-Ready Geospatial Products

Mozhdeh Shahbazi[1]*, Evangelos Bousias Alexakis[1], Mikhail Sokolov[1], Ella Mahoro[1], Pierre Gravel[1]

[1] Canada Centre for Mapping and Earth Observation, Natural Resources Canada, Ottawa, Ontario, Canada, (Mozhdeh.shahbazi; Evangelos.bousiasalexakis; Mikhail.sokolov; Ella.mahoro; Pierre.gravel)@nrcan-rncan.gc.ca
* Corresponding author

Keywords: Historical Air Photos, Feature Matching, Image Alignment, Super-resolution, Colorization, Semantic Segmentation

Abstract

Historical images provide invaluable time series for long-term environmental monitoring and change analysis, particularly in the context of climate-change research. Consequently, growing interest in accessing and exploiting archival airborne imagery has been observed worldwide in recent years. In Canada, the federal government manages the National Air Photo Library, which preserves an extensive collection of historical aerial imagery spanning more than a century. Although this archive is highly valuable, historical photographs in their raw form, whether printed or digitally scanned, are not directly suitable for modern data-driven analytics. Recent advances in computer vision, photogrammetry, and artificial intelligence have created new opportunities to consolidate disparate historical imagery into high-quality digital map products suitable for large-scale automated observations and spatio-temporal analyses. This paper presents solutions applied at Natural Resources Canada for generating analysis-ready mapping products from the National Air Photo Library, focusing on two main workflows: 1) The photogrammetric processing of historical photos with an emphasis on the more challenging automated georeferencing component; 2) Enhancing interpretability through generative artificial intelligence models for super-resolution and deep colorization.

1. Introduction

Knowledge of past conditions plays a vital role in understanding the current state of the environment and anticipating future change. Historical aerial imagery provides a critical long-term record for analyzing environmental and landscape changes and can support applications such as forestation and deforestation monitoring (Bożek et al., 2019; Lydersen and Collins, 2018), urban monitoring and land-cover/land-use analysis (Kupidura et al., 2019; Liu et al., 2018), coastal erosion monitoring and coastal geohazard assessment (Miller et al., 2008), and glacier mass-balance and retreat studies (Fieber et al., 2018; Knuth et al., 2023). By capturing environmental conditions long before the advent of modern satellite programs, these image collections allow researchers to reconstruct past landscapes and develop long-term time-series that are essential for environmental baselining, specifically in the current context of climate change.

The Canadian National Air Photo Library (NAPL) holds over six million historical aerial photographs that span a period of over a century and offer a unique record of how Canada’s landscape has changed over time. This archive represents a major opportunity for producing harmonized, analysis-ready geospatial datasets that support long-term mapping and monitoring workflows. However, transforming these photographs into interoperable digital products at scale remains a significant challenge. A central obstacle in this transformation process is automated georeferencing of historical photos. Particularly, establishing reliable correspondences between historical imagery and modern reference data is hindered by pronounced cross-temporal appearance changes as well as geometric and radiometric differences between historical analog photography and modern digital imagery. This paper outlines solutions being developed at Natural Resources Canada to generate analysis-ready mapping products from NAPL. This includes workflows for photogrammetric processing of historical imagery, with particular attention to the challenges of automated georeferencing, and improving image quality and interpretability using generative Artificial Intelligence (AI) models for super-resolution and deep colorization of historical photos.

Automating the transformation of film-based aerial surveys into orthorectified mosaics and Digital Elevation Models (DEMs) has been an active area of research (Kostrzewa, 2024; Piermattei et al., 2026). Many studies have focused on automating the georeferencing of historical data by identifying Ground Control Points (GCPs) that remain stable across time and can be reliably matched to modern reference imagery that is already tied to a known coordinate system. Given the pronounced temporal, geometric, and radiometric disparities that complicate the reliable identification of cross-epoch correspondences, recent research has increasingly turned to more resilient, often learning-based georeferencing strategies capable of operating under precisely these challenging conditions. The strong performance of learning-based matching methods across challenging scenarios, such as day-to-night (Dusmanu et al., 2019), low-texture (Sun et al., 2021), multi-modal (e.g. radar to optical) (He et al., 2025), and multi-temporal pairs (Morelli et al., 2022), motivated several research studies to explore their suitability for aligning historical imagery to modern satellite or orthorectified mosaics. For instance, to automatically match diachronic images with drastic scene changes, Zhang et al. (2021) developed an approach that first produced historical surface models and then aligned them with contemporary ones using SuperGlue (Sarlin et al., 2020), a graph neural network with attentional message passing that jointly matches correspondences from two sets of features of detected points in each image. Knuth et al. (2023) also proposed co-registering multi-temporal DEMs as a more robust alternative to the more challenging multi-temporal image matching task. However, elevation-based matching methods require high-resolution modern DEMs, which can be challenging to access over a country as vast as Canada. Maiwald et al. (2023) introduced a synergistic strategy that combines SuperGlue-based matches with features derived from a DISK detector (Tyszkiewicz et al., 2020) to improve cross-epoch matching performance. Dahle et al. (2024) developed an automated geo-referencing workflow relying on LightGlue (Lindenberger et al., 2023), a transformer-based deep learning architecture to establish tie-points between historical aerial images of Antarctica and Sentinel-2 satellite images.

Applying colorization techniques can enhance the value of historical aerial images and revitalize archival data by improving the overall interpretability of the data (Farella et al., 2022). Early deep learning methods for automatic colorization relied on end-to-end encoder-decoder architectures based on Convolutional Neural Networks (CNN) that formulated the colorization process either as a regression or classification problem using color-rebalancing strategies to discourage the model from defaulting to dull or desaturated outputs (Iizuka et al., 2016; Zhang et al., 2016). Conditional Generative Adversarial Networks (GAN) further established a general image-to-image framework, routinely applied to colorization amongst other tasks. More recently, diffusion-based approaches such as Control Color (Liang et al., 2025), which support multimodal and interactive image colorization, have demonstrated state-of-the-art colorization quality. In the context of historical aerial imagery, Poterek et al. (2020) applied a conditional GAN for colorizing aerial photographs of the city of Strasbourg. They relied on a gradient penalty scheme called DRAGAN (Kodali et al., 2017) to avoid model collapse, and supplemented the generator objective with a Euclidean-based L1 distance as suggested by Isola et al. (2017). This work demonstrated that conditional GANs can generalize reasonably well to archival aerial surveys, even when trained using synthetically generated grayscale inputs. Farella et al. (2022) introduced Hyper Unet, an encoder-decoder architecture enhanced with Hyper Connections, a type of skip connections inspired by Hyper Columns (Hariharan et al., 2015). Hyper Unet was trained end-to-end using a mean absolute error loss on a training dataset consisting of 10,000 coloured aerial image patches and achieved promising results.

A similar progression can be observed in the development of deep-learning approaches for image super-resolution, which has evolved from early CNN-based models to GAN-driven and, more recently, diffusion-based methods. Initial methods such as SRCNN (Dong et al., 2014) and VDSR (Kim et al., 2016) demonstrated that relatively shallow or moderately deep convolutional architectures could significantly outperform traditional interpolation by learning end-to-end mappings from low-resolution to high-resolution imagery. The introduction of GAN, most notably SRGAN (Ledig et al., 2017) and ESRGAN (Wang et al., 2019), further enhanced perceptual quality by combining adversarial objectives with perceptual losses derived from high-level feature spaces, producing sharper and more photorealistic reconstructions. Recent diffusion-based super-resolution approaches, including pixel-space models such as SR3 (Saharia et al., 2023) and latent-space up-samplers based on Latent Diffusion Models (Rombach et al., 2022), achieve state-of-the-art results by iteratively refining high-frequency details through the reverse diffusion process. These models can be flexibly conditioned not only on the low-resolution input but also on auxiliary guidance signals such as edge maps, semantic segmentation, or structural priors, enabling more controllable and detail-preserving reconstructions (Zhang et al., 2023). Although deep-learning methods for super-resolution have been extensively studied within the broader remote sensing community (Qi et al., 2026; X. Wang et al., 2022), very few have addressed the enhancement of historical aerial imagery. One recent example is the application of GAN-based colorization and super-resolution models to improve rooftop detection performance from historical aerial photographs (Chen et al., 2025).

The rest of the paper is organized as follows. Section 2 outlines the photogrammetric processing pipeline and the colorization and super-resolution workflows. Section 3 presents experimental results for learning-based matching and for generative enhancement models. Section 4 presents the conclusions and future directions for this work.

2. Methodology

2.1 Photogrammetric Processing Pipeline

The first step toward achieving analysis-ready geospatial products is transforming scanned historical photographs into digital mapping products. This process follows a standard photogrammetric workflow, beginning with interior orientation and the identification of matching features across overlapping images. These features are used to establish relative image orientations, which are refined through a self-calibrating bundle adjustment that estimates camera parameters and image orientations in a provisional coordinate system based on approximate flight information. Finally, absolute orientation using automatically extracted GCPs enables the generation of georeferenced point clouds, digital surface models, and ortho-rectified mosaics. In the following paragraphs, only two aspects of interior orientation and absolute orientation, for which in-house solutions are developed, will be discussed. For other typical photogrammetric processes, MicMac software is used (Rupnik et al., 2017).

2.1.1 Restoration of the interior orientation: Historical aerial surveys typically include camera reports or photo engravings that document key interior orientation parameters, such as nominal focal length, physical image dimensions, and fiducial mark configurations. These parameters provide initial estimates for reconstructing camera geometry and establishing a consistent internal reference frame for subsequent bundle adjustments. Fiducial marks play a critical role in estimating the transformation between the scanned image coordinate system and the camera coordinate system. In most photogrammetric software, fiducial mark detection requires manual annotation in the first image of a survey, after which the remaining marks are identified through pattern-matching routines. While generally effective, this approach can be labour intensive and may fail for certain fiducial mark types, particularly simple-shaped marks or degraded markings affected by scan quality, or when even a single fiducial mark is missing. To address these limitations, a deep-learning model is developed and trained to automatically detect and localize fiducial marks. The model architecture combines a Detection Transformer (DETR) model (Carion et al., 2020) chained with a regression module consisting of a Resnet backbone followed by a multi-layer perceptron head. The detector predicts bounding boxes that coarsely localize fiducial marks. The regressor estimates the precise image coordinates of each fiducial mark within its predicted bounding box. To train this model, fiducial marks from a wide range of camera systems represented in the NAPL (Figure 1) are manually annotated on 4750 photos.

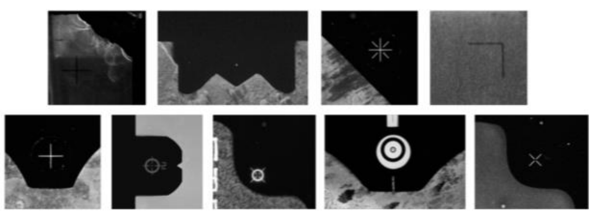


Figure 1. Types of fiducial marks used to train the model.

2.1.2 Absolute Orientation

2.1.2 Absolute Orientation: This step requires reference observations to refine the exterior orientation parameters of the images with respect to a geo-referencing coordinate system. In this work, these reference observations take the form of GCPs obtained by matching historical images to modern satellite imagery. As discussed in Section 1, such cross-temporal matching remains a significant technical challenge due to substantial appearance and geometric differences between historical and modern epochs. To address this challenge, several state-of-the-art learning-based image-matching algorithms with open-source implementations are investigated. The selected methods span a range of matching paradigms, from sparse learned feature approaches to globally conditioned dense correspondence estimators. These learning-based approaches are evaluated against each other, as well as against a classical SIFT-based feature extraction method (Lowe, 2004) combined with a K-Nearest-Neighbour (KNN) matching strategy.

- **D2Net** (Dusmanu et al., 2019): A single CNN which works both as a dense feature descriptor and as a feature detector. Feature detection is performed after feature description, which leads to more stable keypoints compared to detect-then-describe approaches that are based on early detection of shallow/low-level feature representations. The detected features can then be matched using a KNN matcher or any other available matching technique.
- **LightGlue** (LG) (Lindenberger et al., 2023): A transformer-based matcher that is trained to match local features across images. The model has been trained with various feature detectors/descriptors. In this work, three types of features are examined: **ALIKED** (Zhao et al., 2023), a lightweight detector and extractor network introducing a Sparse Deformable Descriptor Head (SDDH) to learn deformable positions of features for each detected keypoint; **SuperPoint** (SP) (DeTone et al., 2018), a CNN model that simultaneously computes pixel-level keypoint locations and the corresponding descriptors in a single pass; **DISK** (Tyszkiewicz et al., 2020), an approach that relies on a Unet-based backbone to produce keypoint heatmaps and dense descriptors while also leveraging reinforcement learning principles.
- **LoFTR** (Sun et al., 2021), **MatchFormer** (Q. Wang et al., 2022), and **COTR** (Jiang et al., 2021): Three approaches that leverage transformers’ global receptive field to learn features that are both context- and position-dependent across the image pairs. To reduce the computational complexity of the transformers, they all rely on coarse-to-fine matching processes.
- **DKM** (Edstedt et al., 2023) and **RoMa** (Edstedt et al., 2024): Two multi-stage approaches that follow the dense feature matching paradigm. Given a pair of images, DKM estimates a dense warp and a dense certainty map from one image to the other. RoMa builds on DKM by incorporating features from a pretrained vision foundation model, which are fused with higher-resolution CNN features to enable precise localization, and by employing a transformer-based decoder that models correspondence uncertainty.

To evaluate these approaches, 60 pairs of historical photographs and corresponding modern georeferenced images are used, spanning multiple acquisition epochs, with the earliest dating back to 1928. The historical photographs, after undergoing the initial self-calibration bundle adjustment step of the photogrammetric workflow, are coarsely georeferenced with positional uncertainties on the order of several hundred meters. Rotational uncertainties are generally lower, typically below 20 degrees, as the flight plans used for initial georeferencing often provide reliable heading information. For each image pair, tie points are manually annotated to define a ground-truth co-registration transformation represented by a homography.

2.2 Historical Image Colorization

AI-based colorization methods are used to enhance the visual appeal and interpretability of grayscale historical orthophotos. In this work, the Pix2Pix GAN architecture (Isola et al., 2017) is adopted. Several network backbones are evaluated as generator architectures, including Hyper Unet (Farella et al., 2022), and Unet with a Resnet-18 encoder, and ColorfulNet (Zhang et al., 2016). The discriminator is a 1x1 PatchGAN, which classifies each pixel independently as real or fake. Additional experiments examine whether colorization is more effective when performed in the transformed Lab colour space or directly in the RGB colour space. Model training relies on modern aerial ortho-photographs acquired from multiple publicly available Canadian datasets. For each color ortho-photo, corresponding NAPL imagery spatially coincident with or proximal to the modern coverage is identified. The modern color images are converted to grayscale and subjected to a series of degradation and domain-adaptation transformations to approximate the radiometric characteristics of historical photographs. These transformations include approximate histogram matching to historical image intensity distributions, as well as the addition of stochastic noise components (Poisson, speckle, and Gaussian) to simulate film grain, scanning artifacts, and radiometric inconsistencies. Using this approach, 25,000 synthetic grayscale-colour image pairs are generated for training, along with 3,750 test pairs, all at a spatial resolution of 512×512 pixels.

2.3 Super-Resolution

Depending on the analog medium available, the scan resolution of historical photographs can be quite limited. For instance, photographs printed on paper are typically scanned at resolutions of only 800-1000 dpi, whereas photographs scanned directly from film rolls can achieve higher resolutions, on the order of 2000-3000 dpi. Beyond limited resolution, historical images often exhibit film grain, blur, scratches, uneven tonal response, and other degradations. As a result, post-enhancement techniques are required to improve the spatial resolution and visual clarity of historical imagery. Deep-learning-based super-resolution and image restoration models offer a promising solution; however, historical photographs often lack high-frequency detail and contain noise characteristics that differ markedly from modern optical imagery, reducing the effectiveness of standard models. To address these limitations, two complementary modeling strategies, deterministic and probabilistic, are investigated. The first approach is based on ESRGAN (Wang et al., 2019), adapted for single-channel grayscale imagery typical of historical aerial datasets. Training is performed using paired low-resolution (LR) and high-resolution (HR) images, allowing the model to learn a direct mapping between the two domains. The second approach is based on a modified Stable Diffusion (SD3) model (Esser et al., 2024). In this framework, the LR input is encoded and injected into the diffusion process alongside the noisy latent representation of the HR image. Unlike ESRGAN, this approach does not rely strictly on pixel-wise supervision; instead, it leverages learned image priors to infer plausible high-frequency details. Consequently, diffusion-based models are better suited for reconstructing globally coherent structures, particularly when the input imagery is heavily degraded.
A critical component of this work is preparing the paired training data. Conventional approaches that synthesize LR images through down-sampling are inadequate, as they fail to reproduce

degradations introduced during low-resolution paper-based scanning. To overcome this limitation, a dedicated paired dataset is created consisting of photographs scanned from paper at 800 dpi and their corresponding film-based scans at 2032 dpi. These image pairs are precisely aligned through automated matching, co-registration, and cropping, after which the HR images are resampled to 1600 dpi to simulate a two-times upscaling factor. The dataset is subsequently tiled into non-overlapping patches of 256×256 (LR) and 512×512 (HR), yielding 149,530 training pairs and 22,430 test pairs.

3. Results and Discussion

3.1 Photogrammetric Processes

The deep learning model developed to tag fiducial marks, as discussed in Section 2.1.1, achieved a success rate of 100% and an average intersection over union of 77.5% in detecting fiducial marks as well as a positioning error of 0.67±0.45 pixels in localizing the centre of the fiducial marks.

To evaluate the automated cross-temporal matching approaches mentioned in Section 2.1.2, two performance metrics are considered: the matching success rate and the mean reprojection error (MRE). The success rate is defined as the proportion of historical-modern image pairs that are successfully matched relative to the total number of pairs in the dataset. The MRE quantifies the discrepancy between the transformation estimated from automatically detected tie points and the ground-truth transformation that aligns each historical image with its corresponding modern counterpart. To compute the MRE, a regular grid of reference points is generated over the historical image plane. These points are projected onto the modern image using both the ground-truth and the automatically estimated transformations, and the resulting differences between corresponding projected coordinates define the reprojection errors. This metric is particularly sensitive to degenerate or poorly distributed matches that may locally fit detected tie points but fail to produce a reliable co-registration over the full image extent. An example of this behaviour is illustrated in Figure 2, where the estimated transformation accurately aligns the top-left region of the image containing the detected keypoints but exhibits large errors elsewhere in the image. Prior to computing the evaluation metrics, an outlier rejection step is applied to the putative correspondences produced by each matching method. Three outlier removal methods are tested: RANSAC and two of its variants, USAC and MAGSAC. Across most experiments, USAC yields the most reliable set of inlier correspondences.

Figure 3 summarizes the matching success rates for all evaluated methods, ordered in descending performance and assessed under three different outlier-removal methods. In these experiments, a match is considered successful if, after outlier rejection, at least nine inlier correspondences remain and the MRE is below 10 pixels. The results indicate that methods following a dense matching paradigm, namely RoMa and DKM, achieve the highest success rates of approximately 90%. Overall, all learning-based matching approaches substantially outperform the traditional SIFT-based matching method.

Figure 4 presents boxplots of the MRE for each matching method. Overall, learning-based approaches achieve comparable, and in many cases lower, MRE values than the traditional SIFT-based method. It should be noted that the statistics shown in Figure 4 correspond only to successfully matched pairs. For example, the SIFT results are computed from approximately 25% of the evaluated pairs, whereas the RoMa results are derived from roughly 90% of the pairs, including a larger proportion of more challenging matching cases.

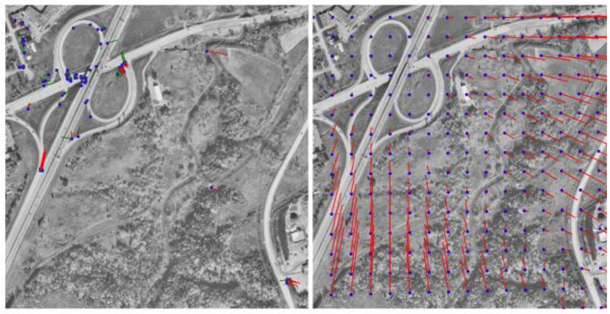


Figure 2. The automatically detected key points and their local matching errors are shown in the left. The reference grid (blue points) and their reprojection errors (red lines) are shown in the right. Errors are magnified tenfold for visualization.

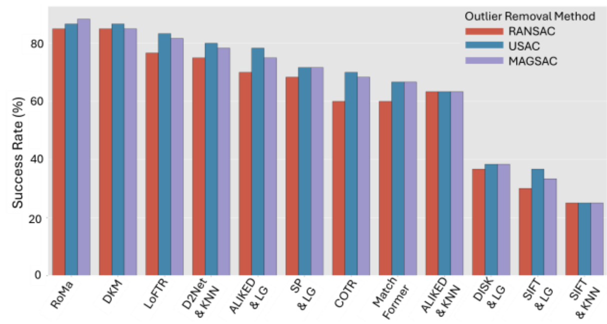


Figure 3. Success rate of the investigated cross-temporal matching methods applied to 60 historical-modern photo pairs.

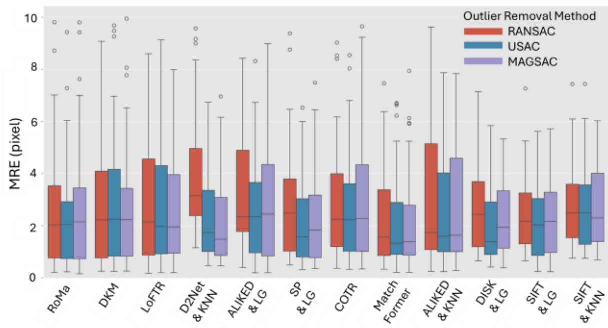


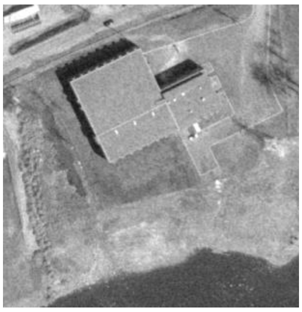
Figure 4. Boxplots of the MRE for each method, computed over successfully matched image pairs.

3.2 Super-Resolution

Two metrics are used to measure the performance of the super-resolution models: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). PSNR quantifies pixel-level reconstruction accuracy, while SSIM assesses perceived structural, brightness, and contrast similarities, both with higher values indicating better performance. As Table 1 indicates, the stable diffusion model outperforms the generative adversarial network (PSNR of 27.62 compared to 24.38). This quantitative improvement is also reflected visually. For instance, in Figure 5, the super-resolved photo by ESRGAN exhibits noticeable artifacts over natural surfaces while the result obtained by SD3 is largely artifact-free and more closely resembles, and in some areas appears even slightly sharper than, the original HR photo.

Model	PSNR	SSIM
ESRGAN	24.38	0.61
SD3	27.62	0.67

Table 1. Performance metrics of ESRGAN and SD3 models for super-resolving (upscaling by a factor of x2) historical photos.


a


b


c


d

Figure 5. Visual output of the super-resolution models. (a) HR photo scanned from film roll at 2032 dpi. (b) LR photo scanned from paper at 800 dpi. (c) LR photo super-resolved from 800 to 1600 dpi using ESRGAN. (d) LR photo super-resolved from 800 to 1600 dpi using SD3.

3.3 Colorization

The quality of colorization models is assessed using pixel-wise, perceptual, and distribution-based metrics. Mean Absolute Error (MAE) quantifies average pixel-level colour errors (the lower, the better). PSNR measures signal fidelity relative to reconstruction noise (the higher, the better). SSIM evaluates the preservation of structural and brightness/contrast information between the colorized and reference images (the higher, the better). DeltaE2000 (D2K) assesses perceptual colour differences, aligning more closely with human visual sensitivity (the lower, the better). Finally, Fréchet Inception Distance (FID) measures the similarity between the distributions of generated and real images in a deep latent space generated by a pretrained Inception-V3 model, capturing higher-level perceptual realism (the lower, the better).

Table 2 summarizes the performance of the developed models. Overall, Hyper Unet in Lab colour space performs the best. As shown in Figure 6 and Figure 7, the outcome of colorization with this model is quite realistic when compared to modern reality.

Once historical photographs are photogrammetrically processed to produce ortho-rectified mosaics and further enhanced through super-resolution and colorization, they can be treated in the same way as modern digital base imagery. Figure 8 illustrates the application of a semantic segmentation model to classify buildings, roads, and forested vegetation in historical imagery. Notably, this model is trained exclusively on modern RGB satellite and aerial data. Applying this pretrained model to super-resolved, colorized historical ortho-photos yields noticeably improved segmentation results compared to direct inference on the original low-resolution grayscale products.

Model	Space	MAE	PS NR	SS IM	D2K	FID
ColorfulNet	Lab	2.34	38.55	0.99	2.77	19.67
ColorfulNet	RGB	8.89	27.08	0.7	4.65	77.61
Hyper Unet	Lab	**2.14**	**39.31**	**0.99**	**2.53**	**13.69**
Hyper Unet	RGB	4.30	34.97	0.99	3.23	14.84
Renset18	Lab	2.20	38.99	0.99	2.61	15.23
Renset18	RGB	4.27	35.33	0.99	3.27	17.23

Table 2. Performance metrics of various Pix2Pix generator backbones for colorizing photos.


a

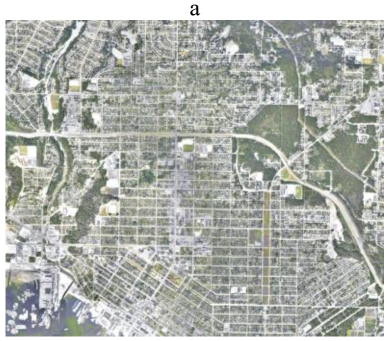

b

Figure 6. Visual output of the colorization model. (a) Grayscale historical ortho-rectified mosaic of an area in British Columbia, Canada, 1967. (b) Historical ortho-rectified mosaic colorized by Hyper Unet Pix2Pix model.


a


b

c

Figure 7. Visual output of the colorization model over a zoomed-in area from Figure 6. (a) Grayscale historical photo. (b) Historical photo colorized by Hyper Unet Pix2Pix model. (c) Modern image of the same area.

a

b

c

Figure 8. Applying semantic segmentation models for classifying buildings (red polygons), roads (pink lines), and forested vegetation (green polygons) on (a) LR historical photo scanned from paper; (b) the same photo artificially super-resolved by a scaling factor of two; (c) the super-resolved photo artificially colorized.

4. Conclusions

This paper presented an integrated set of photogrammetric and deep-learning-based workflows to transform historical aerial photographs from the Canadian National Air Photo Library into analysis-ready geospatial products. By combining advances in automated photogrammetric processing, learning-based cross-temporal image matching, and generative enhancement techniques, the proposed framework enables the large-scale production of georeferenced digital surface models and ortho-rectified mosaics from archival imagery. Experimental results demonstrate that state-of-the-art learning-based matchers, particularly dense correspondence methods, substantially outperform traditional feature-based approaches for automated georeferencing of historical photos. In addition, generative super-resolution and colorization models significantly improve the visual quality and interpretability of historical imagery, enabling the reuse of modern analysis tools such as semantic segmentation models on archival datasets.

While the presented results highlight the strong potential of these methods, several challenges remain and motivate future work. First, further improving the robustness of automated georeferencing under extreme temporal, seasonal, and land-use changes remains a key research direction, particularly for regions with limited stable features. Currently, a feature extractor is being developed based on a foundation model that is pre-trained on both modern RGB images and historical grayscale photos. Incorporating multi-modal reference data, such as vector topographic maps, may further improve alignment reliability. In addition, applying a semantic segmentation model to historical imagery prior to matching may enable more effective filtering and weighting of putative correspondences. For example, matches extracted from forested areas are more likely to be outliers, as vegetation cover changes substantially over long time periods. Second, future work can explore tighter integration of generative enhancement models with photogrammetric pipelines, as opposed to being post-applied to the generated ortho-photos. This, however, necessitates uncertainty-aware super-resolution and colorization strategies that preserve geometric fidelity and avoid introducing misleading artifacts. Third, scaling these workflows to nationwide processing will require continued attention to computational efficiency, human-assisted quality control, and error characterization. Finally, extending downstream analytics, such as land-cover mapping and change detection, into long historical time series will further unlock the scientific value of national aerial archives for environmental monitoring and climate-related studies.

5. References

Bożek, P., Janus, J., Mitka, B., 2019. Analysis of Changes in Forest Structure using Point Clouds from Historical Aerial Photographs. Remote Sensing 11. https://doi.org/10.3390/rs11192259

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-End Object Detection with Transformers. https://doi.org/10.48550/arXiv.2005.12872

Chen, P., Wang, Sicheng, Wang, C., Wang, Senrong, Huang, B., Huang, L., Zang, Z., 2025. A GAN-Enhanced Deep Learning Framework for Rooftop Detection from Historical Aerial Imagery. International Journal of Remote Sensing 46, 6260–6283. https://doi.org/10.1080/01431161.2025.2534994

Dahle, F., Lindenbergh, R., Wouters, B., 2024. Polar perspectives: a deep dive into geo-referencing historical Antarctic photos. International Journal of Digital Earth 17, 2406384. https://doi.org/10.1080/17538947.2024.2406384

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-Supervised Interest Point Detection and Description, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Salt Lake City, UT, USA, pp. 337–33712. https://doi.org/10.1109/CVPRW.2018.00060

Dong, C., Loy, C.C., He, K., Tang, X., 2014. Learning a Deep Convolutional Network for Image Super-Resolution, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 184–199. https://doi.org/10.1007/978-3-319-10593-2_13

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. https://doi.org/10.48550/arXiv.1905.03561

Edstedt, J., Athanasiadis, I., Wadenbäck, M., Felsberg, M., 2023. DKM: Dense Kernelized Feature Matching for Geometry Estimation, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Vancouver, BC, Canada, pp. 17765–17775. https://doi.org/10.1109/CVPR52729.2023.01704

Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M., 2024. RoMa: Robust Dense Feature Matching. IEEE Conference on Computer Vision and Pattern Recognition.

Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., Podell, D., Dockhorn, T., English, Z., Lacey, K., Goodwin, A., Marek, Y., Rombach, R., 2024. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. https://doi.org/10.48550/arXiv.2403.03206

Farella, E.M., Malek, S., Remondino, F., 2022. Colorizing the Past: Deep Learning for the Automatic Colorization of Historical Aerial Images. Journal of Imaging 8. https://doi.org/10.3390/jimaging8100269

Fieber, K.D., Mills, J.P., Miller, P.E., Clarke, L., Ireland, L., Fox, A.J., 2018. Rigorous 3D change determination in Antarctic Peninsula glaciers from stereo WorldView-2 and archival aerial imagery. Remote Sensing of Environment 205, 18–31. https://doi.org/10.1016/j.rse.2017.10.042

Hariharan, B., Arbelaez, P., Girshick, R., Malik, J., 2015. Hypercolumns for object segmentation and fine-grained localization, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 447–456. https://doi.org/10.1109/CVPR.2015.7298642

He, X., Yu, H., Peng, S., Tan, D., Shen, Z., Bao, H., Zhou, X., 2025. MatchAnything: Universal Cross-Modality Image Matching with Large-Scale Pre-Training [WWW Document]. arXiv.org. URL https://arxiv.org/abs/2501.07556v1 (accessed 11.13.25).

Iizuka, S., Simo-Serra, E., Ishikawa, H., 2016. Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Trans. Graph. 35, 110:1-110:11. https://doi.org/10.1145/2897824.2925974

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. CVPR.

Jiang, W., Trulls, E., Hosang, J., Tagliasacchi, A., Yi, K.M., 2021. COTR: Correspondence Transformer for Matching Across Images. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) 6187–6197. https://doi.org/10.1109/ICCV48922.2021.00615

Kim, J., Lee, J.K., Lee, K.M., 2016. Accurate Image Super-Resolution Using Very Deep Convolutional Networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 1646–1654. https://doi.org/10.1109/CVPR.2016.182

Knuth, F., Shean, D., Bhushan, S., Schwat, E., Alexandrov, O., McNeil, C., Dehecq, A., Florentine, C., O’Neel, S., 2023. Historical Structure from Motion (HSfM): Automated processing of historical aerial photographs for long-term topographic change analysis. Remote Sensing of Environment 285, 113379. https://doi.org/10.1016/j.rse.2022.113379

Kodali, N., Abernethy, J., Hays, J., Kira, Z., 2017. On Convergence and Stability of GANs. https://doi.org/10.48550/arXiv.1705.07215

Kostrzewa, A., 2024. Geoprocessing of archival aerial photos and their scientific applications: A review. Reports on Geodesy and Geoinformatics 118, 1–16. https://doi.org/10.2478/rgg-2024-0010

Kupidura, P., Osińska-Skotak, K., Lesisz, K., Podkowa, A., 2019. The Efficacy Analysis of Determining the Wooded and Shrubbed Area Based on Archival Aerial Imagery Using Texture Analysis. ISPRS International Journal of Geo-Information 8. https://doi.org/10.3390/ijgi8100450

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 105–114. https://doi.org/10.1109/CVPR.2017.19

Liang, Z., Li, Z., Zhou, S., Li, C., Loy, C.C., 2025. Control Color: Multimodal Diffusion-Based Interactive Image Colorization. Int J Comput Vis 133, 7897–7923. https://doi.org/10.1007/s11263-025-02549-6

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) 17581–17592. https://doi.org/10.1109/ICCV51070.2023.01616

Liu, D., Toman, E., Fuller, Z., Chen, G., Londo, A., Zhang, X., Zhao, K., 2018. Integration of historical map and aerial imagery to characterize long-term land-use change and landscape dynamics: An object-based analysis via Random Forests. Ecological Indicators 95, 595–605. https://doi.org/10.1016/j.ecolind.2018.08.004

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lydersen, J.M., Collins, B.M., 2018. Change in Vegetation Patterns Over a Large Forested Landscape Based on Historical and Contemporary Aerial Photography. Ecosystems 21, 1348–1363. https://doi.org/10.1007/s10021-018-0225-5

Maiwald, F., Feurer, D., Eltner, A., 2023. Solving photogrammetric cold cases using AI-based image matching: New potential for monitoring the past with historical aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 206, 184–200. https://doi.org/10.1016/j.isprsjprs.2023.11.008

Miller, P., Mills, J., Edwards, S., Bryan, P., Marsh, S., Mitchell, H., Hobbs, P., 2008. A robust surface matching technique for coastal geohazard assessment and management. ISPRS Journal of Photogrammetry and Remote Sensing, Theme Issue: Remote Sensing of the Coastal Ecosystems 63, 529–542. https://doi.org/10.1016/j.isprsjprs.2008.02.003

Morelli, L., Bellavia, F., Menna, F., Remondino, F., 2022. Photogrammetry Now and Then – From Hand-Crafted to Deep-Learning Tie Points –. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2-W1-2022, 163–170. https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-163-2022

Piermattei, L., McNabb, R., Elias, M., Ressl, C., Dehecq, A., Girod, L., Dewez, T., Eltner, A., 2026. Unlocking the Past: A review of digital processing of historical aerial and satellite stereo analog imagery for geoscience applications. IEEE Geosci. Remote Sens. Mag. 2–33. https://doi.org/10.1109/MGRS.2025.3645144

Poterek, Q., Herrault, P.-A., Skupinski, G., Sheeren, D., 2020. Deep Learning for Automatic Colorization of Legacy Grayscale Aerial Photographs. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 2899–2915. https://doi.org/10.1109/JSTARS.2020.2992082

Qi, Y., Lou, M., Liu, Y., Li, L., Yang, Z., Nie, W., 2026. Advancing image super-resolution techniques in remote sensing: A comprehensive survey. ISPRS Journal of Photogrammetry and Remote Sensing 231, 68–100. https://doi.org/10.1016/j.isprsjprs.2025.10.024

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis with Latent Diffusion Models, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, pp. 10674–10685. https://doi.org/10.1109/cvpr52688.2022.01042

Rupnik, E., Daakir, M., Pierrot Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. Open geospatial data, softw. stand. 2, 14. https://doi.org/10.1186/s40965-017-0027-2

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M., 2023. Image Super-Resolution via Iterative Refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 4713–4726. https://doi.org/10.1109/TPAMI.2022.3204461

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning Feature Matching With Graph Neural Networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 4937–4946. https://doi.org/10.1109/CVPR42600.2020.00499

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-Free Local Feature Matching with Transformers, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 8918–8927. https://doi.org/10.1109/CVPR46437.2021.00881

Tyszkiewicz, M., Fua, P., Trulls, E., 2020. DISK: Learning local features with policy gradient. ArXiv.

Wang, Q., Zhang, J., Yang, K., Peng, K., Stiefelhagen, R., 2022. MatchFormer: Interleaving Attention in Transformers for Feature Matching.

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., Min, H., 2022. A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing. Remote Sensing 14. https://doi.org/10.3390/rs14215423

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C., 2019. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, in: Leal-Taixé, L., Roth, S. (Eds.), Computer Vision – ECCV 2018 Workshops, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 63–79. https://doi.org/10.1007/978-3-030-11021-5_5

Zhang, L., Rao, A., Agrawala, M., 2023. Adding Conditional Control to Text-to-Image Diffusion Models. https://doi.org/10.48550/arXiv.2302.05543

Zhang, L., Rupnik, E., Pierrot-Deseilligny, M., 2021. Feature matching for multi-epoch historical aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 182, 176–189. https://doi.org/10.1016/j.isprsjprs.2021.10.008

Zhang, R., Isola, P., Efros, A.A., 2016. Colorful Image Colorization, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 649–666. https://doi.org/10.1007/978-3-319-46487-9_40

Zhao, X., Wu, X., Chen, W., Chen, P.C.Y., Xu, Q., Li, Z., 2023. ALIKED: A Lighter Keypoint and Descriptor Extraction Network via Deformable Transformation.