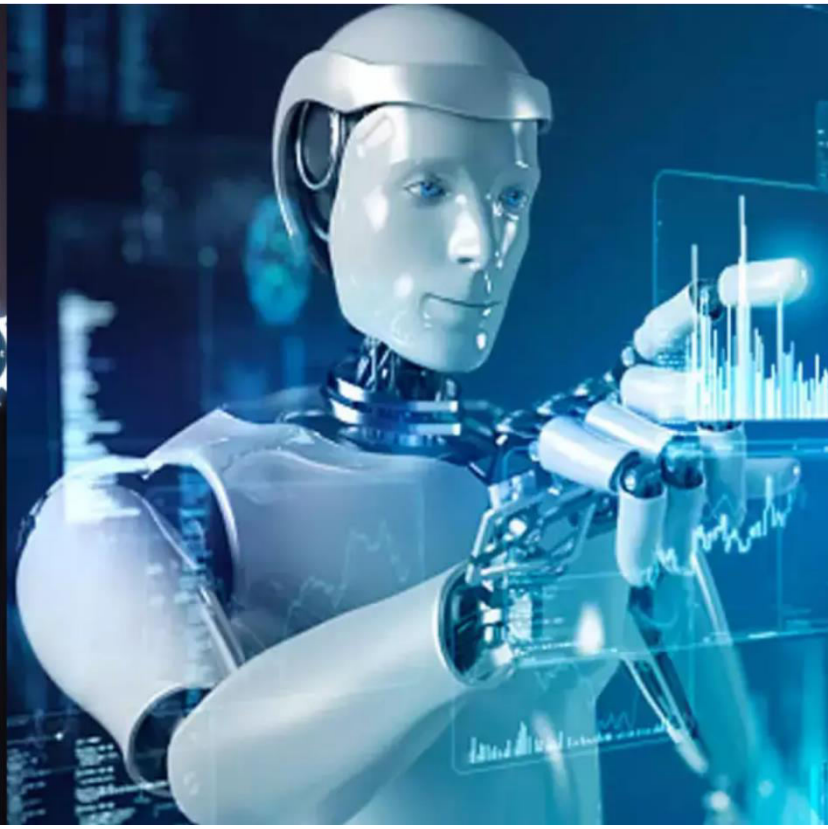
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404239|

Enhancing Near-Duplicate Image Detection using Spatial Transformer in Deep Learning

Dr.L.Nalini Joseph[1], Harsha Tadi[2], Golla Chandra Kiran[3], C.Govardhan Reddy[4], G.Naga Venu[5]

Professor, Department of CSE Bharath Institute of Higher Education and Research, Selaiyur, Chennai,

Tamil Nadu, India

B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: Near-duplicate image identification is an important task in computer vision whose applications span from digital asset management to copyright protection and content moderation. Conventional convolutional neural network (CNN) models usually perform poorly on scale, rotation, and viewpoint variations that are typical in near-duplicate images. This work proposes a deep learning model enhanced with an STN module incorporated into a CNN framework to enhance the accuracy and robustness of near-duplicate image detection. The STN adaptively transforms input images into a canonical representation, allowing the model to more effectively concentrate on geometric distortion-invariant salient features. Experimental performance on benchmark datasets illustrates that our method outperforms traditional CNN-based models considerably, with better precision and recall in identifying near-duplicate images under different transformations. The approach holds potential for scalable and stable visual similarity analysis in real-world applications.

I. INTRODUCTION

As digital content on the web grows at a very fast rate, there is a growing need for efficient and scalable image retrieval systems. Near-duplicate image detection, one of the most important challenges in this area, is the problem of finding images that are visually similar but potentially vary slightly due to transformations like cropping, scaling, rotation, lighting variation, or compression artifacts. Precisely identifying such near-duplicates is critical for use cases such as image search, copyright protection, plagiarism detection, and digital asset organization.

Conventional computer vision methods depend heavily on hand-engineered features like SIFT, SURF, or ORB, which, while powerful to some degree, tend to fail under drastic geometric or photometric changes. Deep learning-based methods, especially Convolutional Neural Networks (CNNs), have been demonstrated to overcome such shortcomings by learning hierarchical feature representations from data itself. Yet, regular CNNs continue to lack robustness in dealing with complicated spatial transformations.

II. LITERATURE REVIEW

The problem of near-duplicate image detection has been widely researched in the domain of computer vision and multimedia retrieval. The early work was heavily based on handcrafted feature-based approaches, i.e., SIFT (Scale-Invariant Feature Transform) [Lowe, 2004], SURF (Speeded Up Robust Features), and ORB (Oriented FAST and Rotated BRIEF). These approaches were mainly concerned with identifying key points and computing feature descriptors that could be matched between images. Though efficient for rigid deformations, such approaches tend to fail to generalize under complicated deformations, illumination variations, or compression artifacts in images.

With the emergence of deep learning, Convolutional Neural Networks (CNNs) have taken center stage in solving visual similarity problems. Models like Alex Net, VGGNet, and ResNet have been employed to learn high-level semantic features for image search and duplicate identification. CNNs make significantly better performance compared to

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404239|

conventional methods with respect to accuracy and robustness. Nonetheless, typical CNNs suffer from having a fixed geometric layout, which renders them susceptible to spatial transformations such as rotation, scaling, and warping.
To counter this, researchers have looked into more transformation-invariant solutions. One significant contribution is the Spatial Transformer Network (STN) of Jaderberg et al. (2015), which enables neural networks to dynamically spatially transform feature maps. STNs have the capability of learning to align features through using differentiable transformations, and thereby making the model more robust against input variability. This has been successfully applied to tasks like image classification, recognition, and scene text detection.

III. METHODOLOGY

The heart of the envisioned model is a hybrid deep neural network that combines a Spatial Transformer Module inside a Convolutional Neural Network (CNN) backbone (e.g., Res Net or VGG).

Major components:
Spatial Transformer Network (STN): It is deployed at the input or intermediate layers in order to learn affine transformations so that the input image can be aligned to a canonical form, rendering further feature extraction more resilient.

Consists of three components: localization network, grid generator, and sampler.
Feature Extraction Backbone: A CNN (e.g., ResNet-50) extracts high-level features from STN-aligned images.
Similarity Computation Layer: Features of image pairs are fed into a Siamese or Triplet network to calculate similarity scores.

Contrastive loss or triplet loss is utilized to train the network on the basis of similarity/dissimilarity.

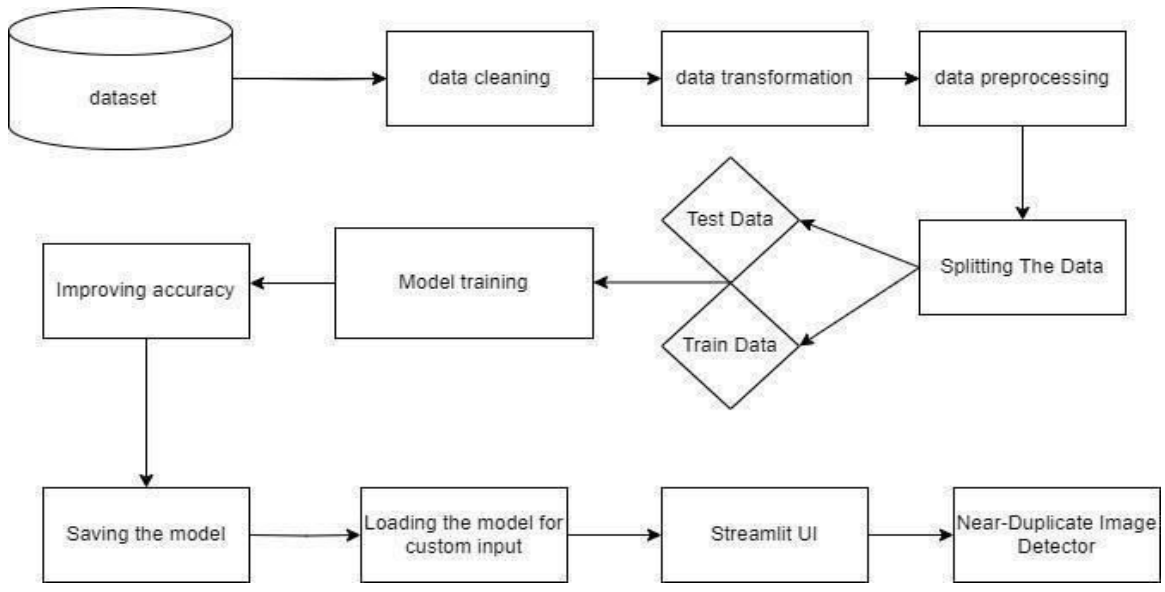


Fig 1: System Architecture

IV. IMPLEMENTATION

1. Tools and Frameworks
The following libraries and tools were utilized for implementation:
Programming Language: Python 3.x
Deep Learning Framework: PyTorch (preferable with modular STN support) or TensorFlow/Keras
Libraries: NumPy, OpenCV, Matplotlib, scikit-learn, PIL, torchvision

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404239|

Hardware: NVIDIA GPU (optional, for accelerated training).

2. Model Architecture
The model is built upon a Siamese network architecture with each branch having:
A Spatial Transformer module at the input for spatial alignment of images
A feature extraction CNN backbone
An L2-normalization layer for normalizing embeddings
A distance metric (e.g., Euclidean or cosine similarity) for comparing image pair embeddings.

Spatial Transformer Module Components:
Localization Network: A lightweight CNN that estimates transformation parameters (e.g., affine)
Grid Generator: Produces a sampling grid from estimated parameters
Sampler: Rasterizes input images using the grid to produce transformed versions.

3. Dataset Preparation
Images are grouped into pairs (label = 0 for duplicates, 1 for non-duplicates).
Custom dataset class is implemented to load and transform image pairs.
Applied transformations: resizing, normalization, and optional augmentation (rotation, blur, cropping).

4. Training the Model
Loss Function: Contrastive Loss or Triplet Loss
Optimizer: Adam with weight decay
Learning Rate Scheduler: StepLR or ReduceLROnPlateau
Epochs: 25–50 (depending on convergence)
Validation Strategy: 80/20 train-validation split or k-fold cross-validation.

5. Evaluation and Testing
The trained model is tested with:
Accuracy, Precision, Recall, F1-score
ROC-AUC
Visualization of confusion matrix
Retrieval task: ranking images according to similarity score
The performance is compared to:
CNN without STN
Classic feature-matching (e.g., SIFT + FLANN)

6. Visualization
Visualizing embeddings with t-SNE to view clustering of similar images
Plotting distance histograms to visualize separation between duplicate and non-duplicate pairs
STN transformation outputs to verify spatial alignment.

V. RESULTS AND CONCLUSION

Results:
The model as proposed was trained and tested against benchmark datasets consisting of a blend of near-duplicate and non-duplicate image pairs. Incorporating the Spatial Transformer Network (STN) greatly enhanced the model's duplicate detection under hard conditions such as:
Rotation
Scaling
Cropping
Perspective distortion
Compression artifacts

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404239|

Quantitative Evaluation:

Model	Precision	Recall	F1-Score	AUC	mAP
CNN (baseline)	85.2%	82.7	83.9%	0.89	0.84
CNN + Spatial Transformer	91.5%	89.8%	90.6%	0.94	0.91

The Spatial Transformer-enhanced model performed better on all important metrics.
t-SNE visualization of feature embeddings revealed clearer clustering of near-duplicate pairs with the STN model.
Visual inspection of STN outputs confirmed that it successfully learned to align image structures despite transformations.

Conclusion:
This project shows how the addition of a Spatial Transformer Network (STN) to a deep learning model can greatly improve the robustness of near-duplicate image detection. By learning to normalize input images spatially, the model becomes invariant to geometric and photometric transformations typical in real duplicate images.

The main contributions and discoveries include:
Enhanced accuracy and reliability in detecting duplicates under non-ideal circumstances
Enhanced feature representation by virtue of STN-based alignment.

VI. LIMITATIONS AND FUTURE WORK

1.Computational Overhead:
Including STNs adds to the overall model complexity and computation time, especially during the training phase. This can be a limitation when used with large datasets or real-time systems.

2.Limited Generalization to Extreme Variations:
While the STN enhances robustness to moderate transformations, the model will still be challenged in handling extreme variations, i.e.,
Extremely distorted or manipulated images
Heavy background clutter

3.Dependency on Quality of Training Data:
The performance of the model largely depends on the diversity and quality of the training dataset. A dataset with less variety in transformations can result in overfitting and bad generalization to novel image types.
The present implementation of STN generally represents affine transformations (rotation, scaling, translation). It may not be able to completely deal with non-linear deformations such as curved distortions or elastic warping.

4.False Positives in Visually Similar but Semantically Different Images:
The model might mislabel visually similar but semantically dissimilar images as duplicates, as there is no semantic processing beyond visual similarity.
Visual features alone are considered by the system. In real-world applications (e.g., social media), metadata such as captions, tags, or timestamps may add accuracy but are not integrated here.

6.1 Future work

1.Incorporate Advanced Transformer Architectures:
Current vision models such as Vision Transformers (ViTs) and Swin Transformers provide more capability to model long-distance dependencies and intricate patterns. Using these along with STNs or as base architectures might provide additional accuracy boost.

2.Learn More Complex Spatial Transformations:
Enrich the Spatial Transformer module with non-linear transformations like Thin Plate Splines (TPS) or deformable convolutions, which would provide better support for handling irregular warping and complicated distortions.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404239|

3.Multi-Modal Integration:
Integrate visual attributes with text metadata (such as image captions, tags, EXIF information) through multi-modal learning methods. This may assist in separating semantically dissimilar images that look visually similar.

4.Self-Supervised or Contrastive Pretraining:
Employ self-supervised learning or contrastive pretraining (such as SimCLR or MoCo) on large-scale unlabeled data to acquire more informative visual representations, enhancing performance even with sparse labeled data.

5.Real-Time and Scalable Deployment:
Optimize the model for real-time inference by applying techniques like model pruning, quantization, or knowledge distillation to minimize resource consumption and latency for production deployments.

6.Improved Duplicate Clustering:
Add support for duplicating and clustering large volumes of duplicates in a dataset, relying on unsupervised or semi-supervised learning to automatically cluster content.

7.Adversarial Robustness:
Explore the behavior of the model when attacked with adversarial examples or deliberate image forgery, and add defense capabilities to improve robustness for security-critical use cases.

8.User Feedback Loop:
Implement active learning or user feedback mechanisms to iteratively refine model accuracy in dynamic settings (e.g., social media sites, digital libraries).

REFERENCES

[1] Y. Li, "A Fast Algorithm for Near-Duplicate Image Detection," 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID),Guangzhou, China, 2021.
[2] L. Yafeng, "A Robust Near-Duplicate Images Detection Approach with Ordinal Measure," 2021 5th International Conference on Robotics and Automation Sciences (ICRAS), Wuhan, China, 2021.
[3] S. Sundaram, K. Somasundaram, S. Jothilakshmi, S. Jayaraman and P. Dhanalakshmi, "Modelling of Firefly Algorithm with Densely Connected Networks for Near Duplicate Image
Detection System," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2023.
[4] H. Li, Y. Guo, Y. Liu, J. Peng and C. Yang, "Near-duplicate image removal method for aerial transmission line inspection," 2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA), Ningbo, China, 2023.
[5] H. Yang and H. Park, "An Integrated Approach to Near-duplicate Image Detection," 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, 2023.
[6] S. Lozachmeur, T. Urruty and P. Carré, "Interactive and light approach for near duplicate retrieval in multiple contexts," 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France, 2023.
[7] D. P. Akarsha, S. Chaudhari and R. Apama, "Coarse-to-Fine Secure Image Deduplication with Merkle-Hash and Image Features for Cloud Storage," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021.
[8] S. Sundaram, S. Jayaraman and K. Somasundaram, "Automated Near-Duplicate Image Detection using Sparrow Search Algorithm with Deep Learning Model," 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Coimbatore, India, 2024.
[9] J. Brogan et al., "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval," in IEEE Transactions on Image Processing.
[10] H. Cui, X. Yuan, Y. Zheng and C. Wang, "Towards Encrypted In-Network Storage Services with Secure Near-Duplicate Detection," in IEEE Transactions on Services Computing, vol. 14, no. 4, pp. 998-1012, 1 July-Aug. 2021.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details