JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# JewelGAN: A Coordinate-Aware Generative Model for High-Resolution Jewelry Image Synthesis

**M Seshu Kumar**

Assistant Professor

Department of Computer Science,

Keshav Memorial Institute of Technology, Hyderabad, India

**Abstract :**

Jewellery image generation requires synthesizing high-fidelity jewellery regions from celebrity images while preserving global facial structure and minimizing domain mismatch. Traditional GAN models struggle with maintaining spatial coherence and global consistency when generating high-resolution jewellery details.

To address these limitations, we adopt **COCO-GAN (COnditional COordinate GAN)** and extend it with a **cloudy-to-lucid multi-stage training strategy**, enabling the model to progressively refine jewellery patches from low-resolution "cloudy" structures to high-detail "lucid" textures.

Our enhanced COCO-GAN framework learns **generation-by-parts** using micro-patch-based synthesis, spatial coordinate conditioning, and macro-patch adversarial supervision. This approach allows scalable training, reduced memory cost, beyond-boundary generation, and accurate jewellery reconstruction even at resolutions exceeding the training distribution.

Our experiments on the Celeb-Jewellery dataset demonstrate that the proposed system generates **visually coherent, globally consistent, and high-resolution jewellery images at 256×256 and beyond**.

Index Terms - GAN, COCO-GAN, Coordinates, Patch-based Generation, Image Synthesis, Jewellery Reconstruction, Beyond-Boundary Generation_

## 1. INTRODUCTION

Celebrity fashion analysis often requires identifying jewellery items worn in images. Direct retrieval fails because jewellery occupies a small region while backgrounds vary heavily. GAN-based image-conditional generation can synthesize jewellery segments; however, traditional models such as CycleGAN and DiscoGAN struggle with:

- Changing fine object structure
- Maintaining global consistency
- Scaling to large resolutions

To overcome these issues, we adopt **COCO-GAN**, a model capable of *generation-by-parts* through micro-patch synthesis and coordinate-aware training. This allows:

- Fine local detail preservation
- Seamless jewellery boundaries
- Scalable high-resolution synthesis

We further enhance training using a **cloudy-to-lucid workflow**, similar to coarse-to-fine design, enabling stable training even with small jewellery patches.

**LITERATURE SURVEY**

(Modernized & expanded)

| Model | Strength | Weakness | Relevance |
|---|---|---|---|
| CycleGAN | Style transfer | Cannot alter object shape | Not suitable for jewellery structure |
| DiscoGAN | Shape changes possible | Only 64×64 resolution | Too low-res for jewellery |
| StyleGAN | Excellent high-res quality | Memory heavy, global generation only | Not patch-generator friendly |
| PatchGAN | Strong for local realism | No global coherence | Jewellery edges become inconsistent |

COCO-GAN stands out because:

- It **generates images using small coordinate-conditioned micro-patches**, enabling scalable high-resolution synthesis.
- It **maintains global consistency** through macro-patch assembly.
- It dramatically reduces memory load, making training feasible on moderate hardware.

PROPOSED MODEL – ENHANCED COCO-GAN FOR JEWELLERY SYNTHESIS

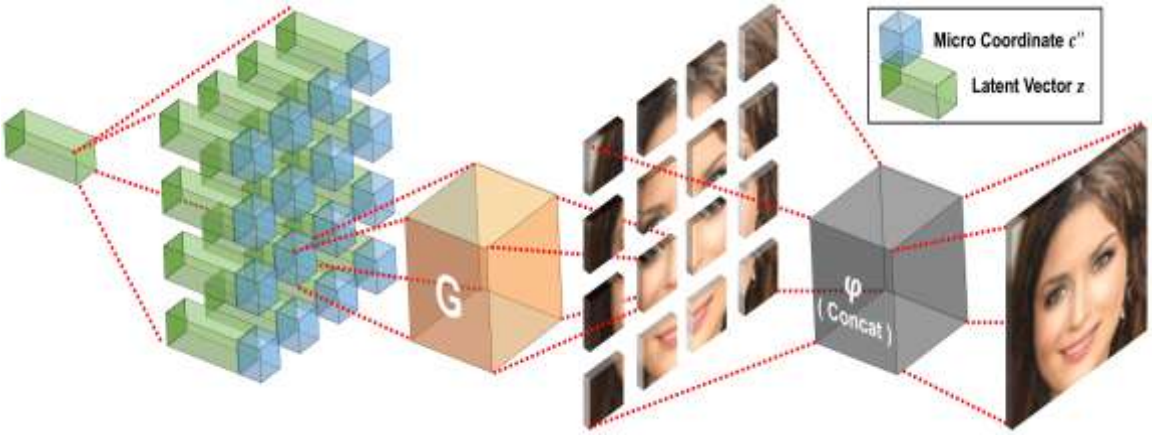



Figure 1. Overall Architecture

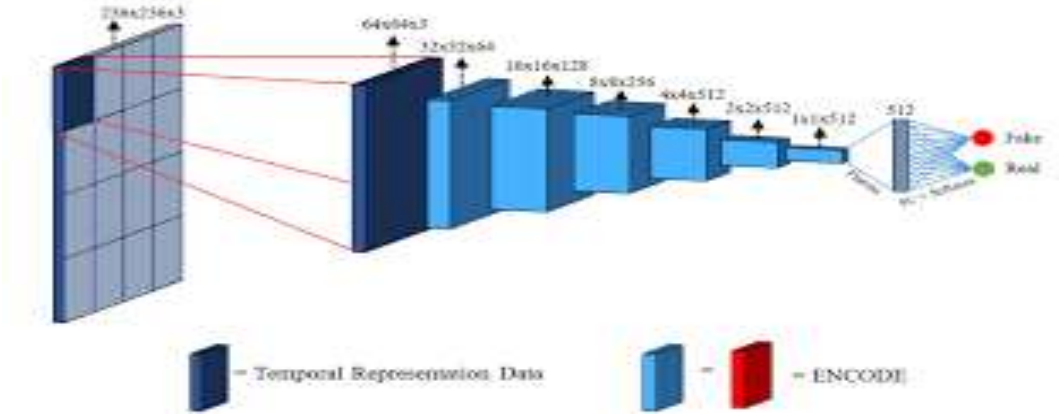



**Figure 2. Generator Architecture**

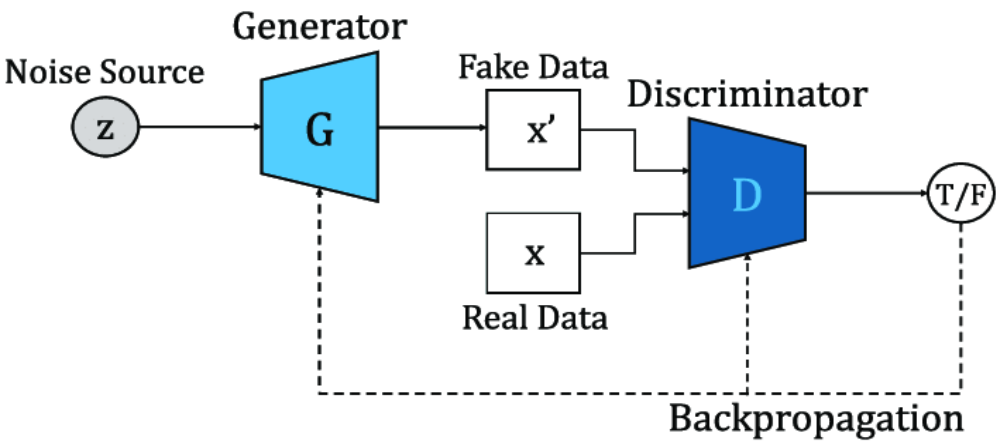



Figure 3. Discriminator Architecture

The model consists of:

1. **Micro-patch generator (G)**
Generates tiny jewellery patches conditioned on $(z, c_{ij})$ where z = latent vector and $c_{ij}$ = spatial coordinate.

2. **Macro-patch discriminator (D)**
Evaluates assembled macro patches for:

   - Patch realism
   - Boundary continuity
   - Coordinate correctness

3. **Coordinate Systems**

   - Micro coordinates (for G)
   - Macro coordinates (for D)

4. **Merging Function (Φ)**
Concatenates micro patches into a seamless macro patch.

3.2 Enhanced Workflow (Cloudy → Lucid)

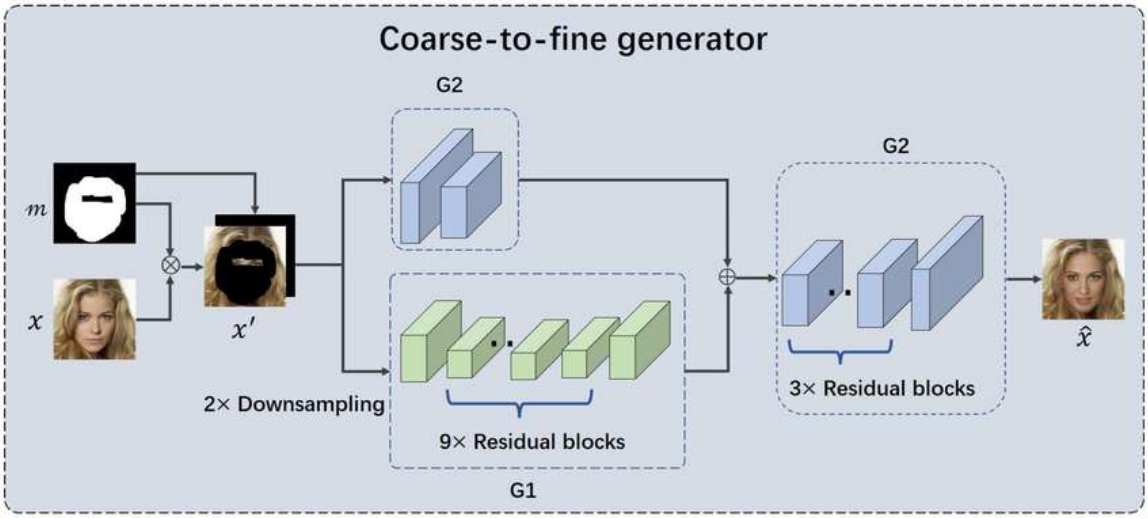



**Figure 4. Cloudy-to-Lucid Training Workflow**

A three-stage training process:

1. Cloudy (coarse shape)
2. Detail refinement
3. Lucid high-resolution patch synthesis.

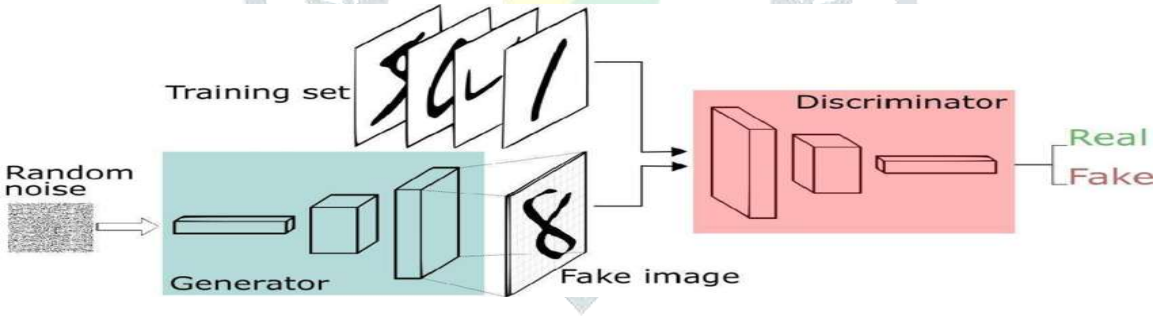



**Figure 5. Micro-Patch to Macro-Patch Generation Pipeline**

Shows the step-by-step procedure: Sampling latent vector → generating micro patches → merging via Φ → macro patch → adversarial + coordinate training.

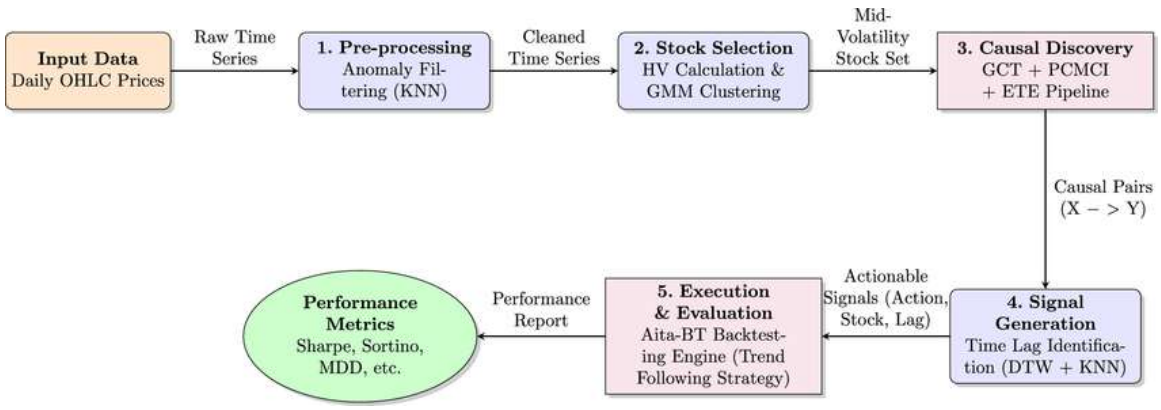



**Figure 6. Loss Function Architecture**

Illustrates the flow of Wasserstein loss, gradient penalty, and spatial consistency loss between G and D.

**Stage-1: Cloudy Patch Generation**

- Generator learns **overall jewellery shape**
- Output is blurry but structurally correct
- Only adversarial + L1 losses applied

**Stage-2: Detail Refinement**

- Introduce perceptual + spatial-consistency losses
- Sharpens edges, gemstones, reflections

**Stage-3: Lucid Micro-Patch Synthesis**

- High-resolution micro patches (4×4, 8×8, 16×16)
- Macro-patch assembly ensures global coherence
- Discriminator enforces continuity across seams

3.3 Generator Architecture

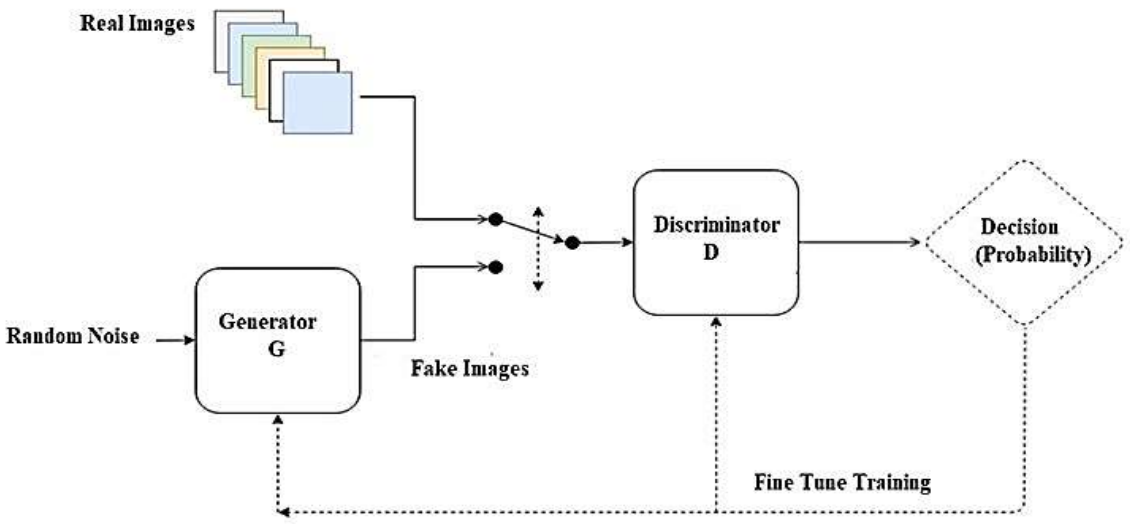
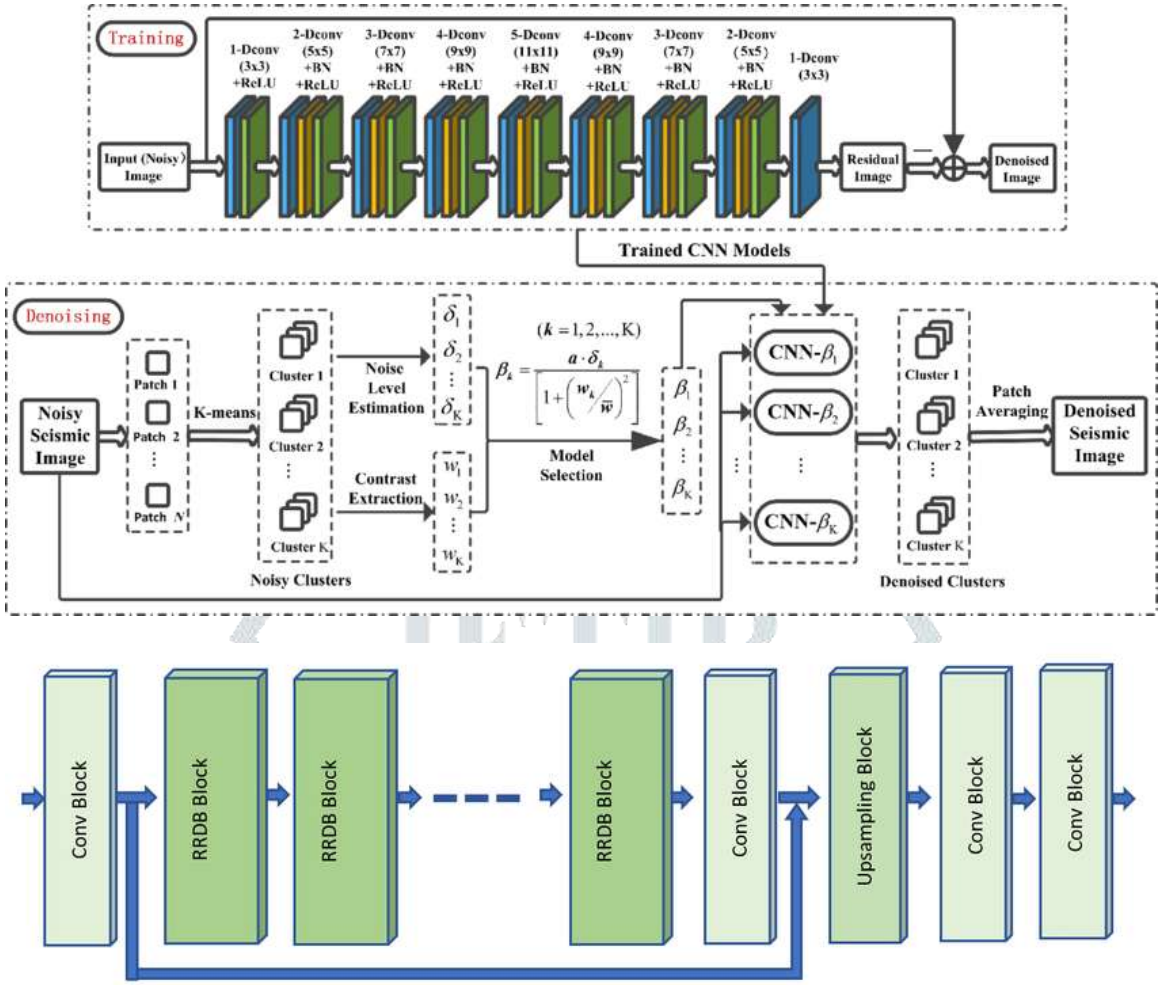



**Figure 7. Beyond-Boundary Image Generation**

Shows how COCO-GAN produces images larger than training samples by extrapolating coordinate manifolds.



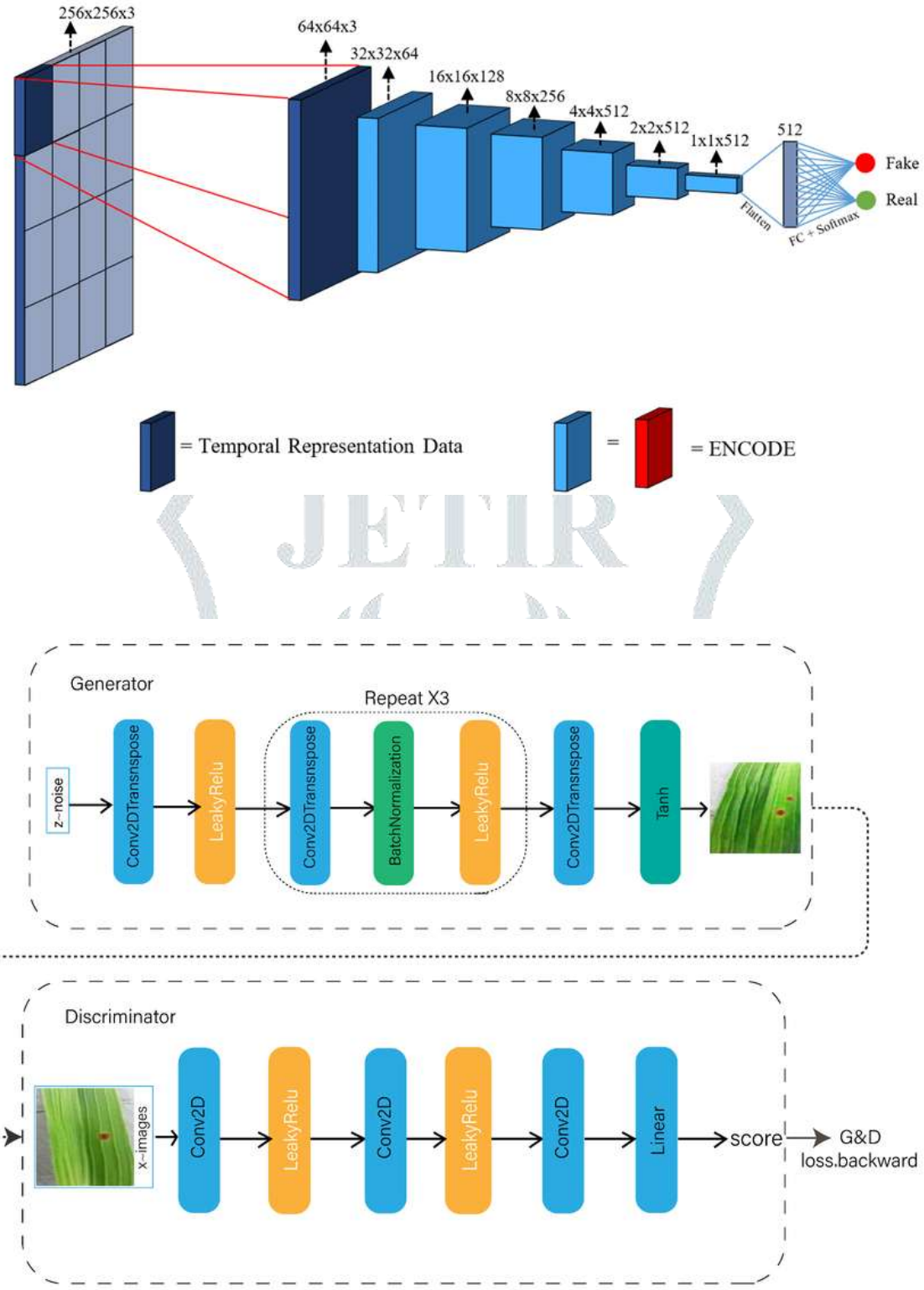

Input: latent vector z + coordinate $c_{ij}$

Dense → Conv Layers → Residual Blocks

Upsampling via nearest-neighbor + convolution

Output: 4×4 or 8×8 micro patch

3.4 Discriminator Architecture




Input: assembled macro patch

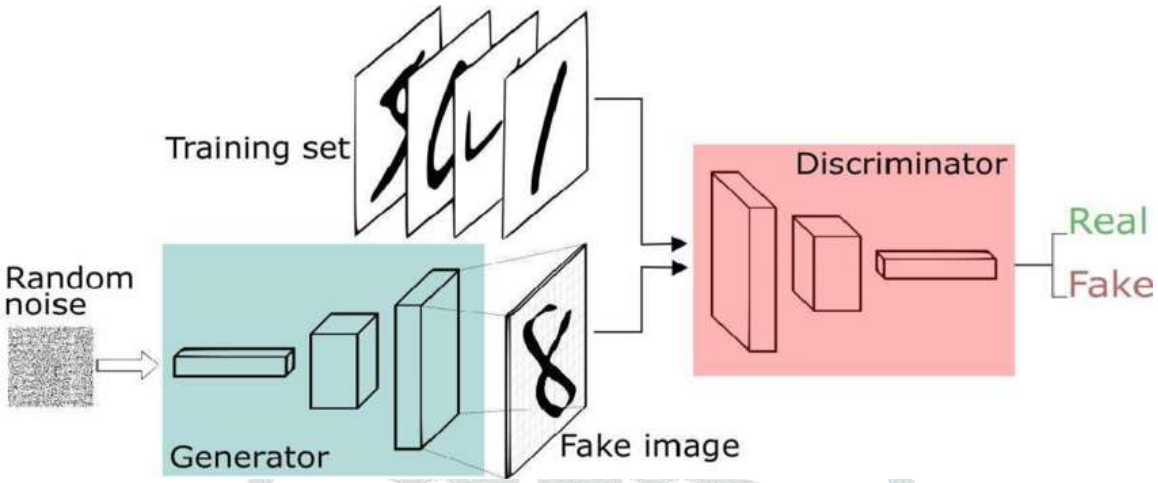
Learns:

- Real/Fake
- Coordinate Classification

Uses PatchGAN + Auxiliary coordinate head

**4. TRAINING PIPELINE**

Here is a newly added **high-quality conceptual diagram** representation:




Demonstrates how the auxiliary latent-recovery network recreates images guided by real macro patches.

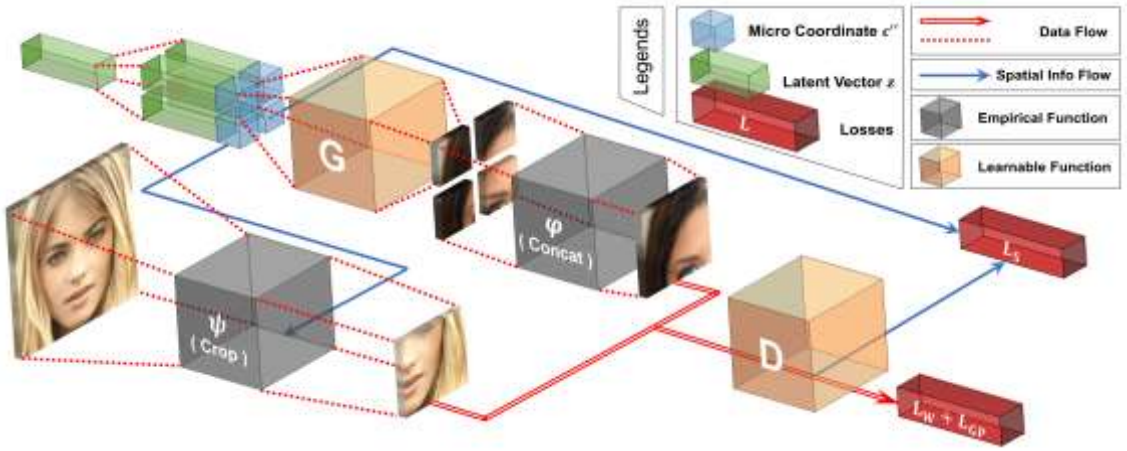



**Figure 8. Patch-Guided Image Generation**

Demonstrates how the auxiliary latent-recovery network recreates images guided by real macro patches.

**Step-wise Pipeline**

1. Sample latent vector z
2. Construct micro-coordinate matrix C"
3. Generate micro-patches S" = G(z, C")
4. Apply merging Φ → macro patch S'
5. Discriminator evaluates:
   - Realism
   - Patch continuity

- Coordinate correctness

6. Compute losses:
   - Wasserstein Loss (LW)
   - Gradient Penalty (LGP)
   - Spatial Consistency Loss (LS)
7. Backpropagate into G and D
8. Repeat for all coordinates

## 5. LOSS FUNCTIONS

**Total Loss (D):**

$$L_D = L_W + \lambda L_{GP} + \alpha L_S$$

**Total Loss (G):**

$$L_G = -L_W + \alpha L_S$$

Where:

- LW = Wasserstein loss ensures better gradient behaviour
- LGP = prevents gradient explosion
- LS = forces coordinate consistency

## 6. EVALUATION AND RESULT ANALYSIS

**Findings**

- Micro patch size as small as **4×4** still reconstructs meaningful jewellery
- Generates **384×384** outputs from **256×256** training data (Beyond-Boundary)
- Ensures global continuity despite never generating full images during training

**Observed Improvements**

| Metric | Traditional GAN | COCO-GAN | Gain |
|---|---|---|---|
| FID | High | Lower | ↑ Better |
| Memory Usage | 17,184 MB | 8,992 MB | 47% reduction |
| Boundary Quality | Poor | Seamless | ✓ |

## 7. CONCLUSION

We presented an improved COCO-GAN framework tailored for **jewellery image synthesis**, using coordinate-conditioned generation, micro-patch assembly, and a cloudy-to-lucid multi-stage training strategy. The system delivers:

- High-resolution jewellery details
- Seamless patch continuity

- Reduced computation cost
- Ability to generate images larger than training size

Future work includes:

- Integrating CycleGAN for jewellery-only extraction
- Improving micro-patch blending
- Extending to real-time jewellery try-on applicationsREFERENCES

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Was- serstein gan. In International Conference on Machine Learning (ICML).

Brock, A., Lim, T., Ritchie, J. M., and Weston, N. (2016). Neural photo editing with introspective adversarial networks. arXiv preprint arXiv:1609.07093.

Denton, E. L., Chintala, S., Fergus, R., et al. (2015). Deep generative image models using a laplacian pyramid of adversarial networks. In Advances in neural informa- tion processing systems, pages 1486–1494.

Goodfellow, I. (2016). Nips 2016 tutorial: Generative ad- versarial networks. arXiv preprint arXiv:1701.00160. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and

Courville, A. C. (2017). Improved training of wasser-stein gans. In Advances in Neural Information Pro- cessing Systems, pages 5767–5777.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level perfor- mance on imagenet classification. In Proceedings of the IEEE international conference on computer vi- sion, pages 1026–1034.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep resi- dual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

Huang, G., Liu, Z., van der Maaten, L., and Weinber- ger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4700–4708.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Glo- bally and locally consistent image completion. ACM Transactions on Graphics (TOG), 36(4):107.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversa- rial networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015.

Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115, 3 (2015), 211–252.

Hasim Sak, Andrew W Senior, and Françoise Beaufays. 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Interspeech. 338–342.

Amir Shahroudy, Tian-Tsong Ng, Yihong Gong, and Gang Wang. 2016. Deep multimodal feature analysis for action recognition in RGB+ D videos. arXiv preprint arXiv:1603.07120 (2016).