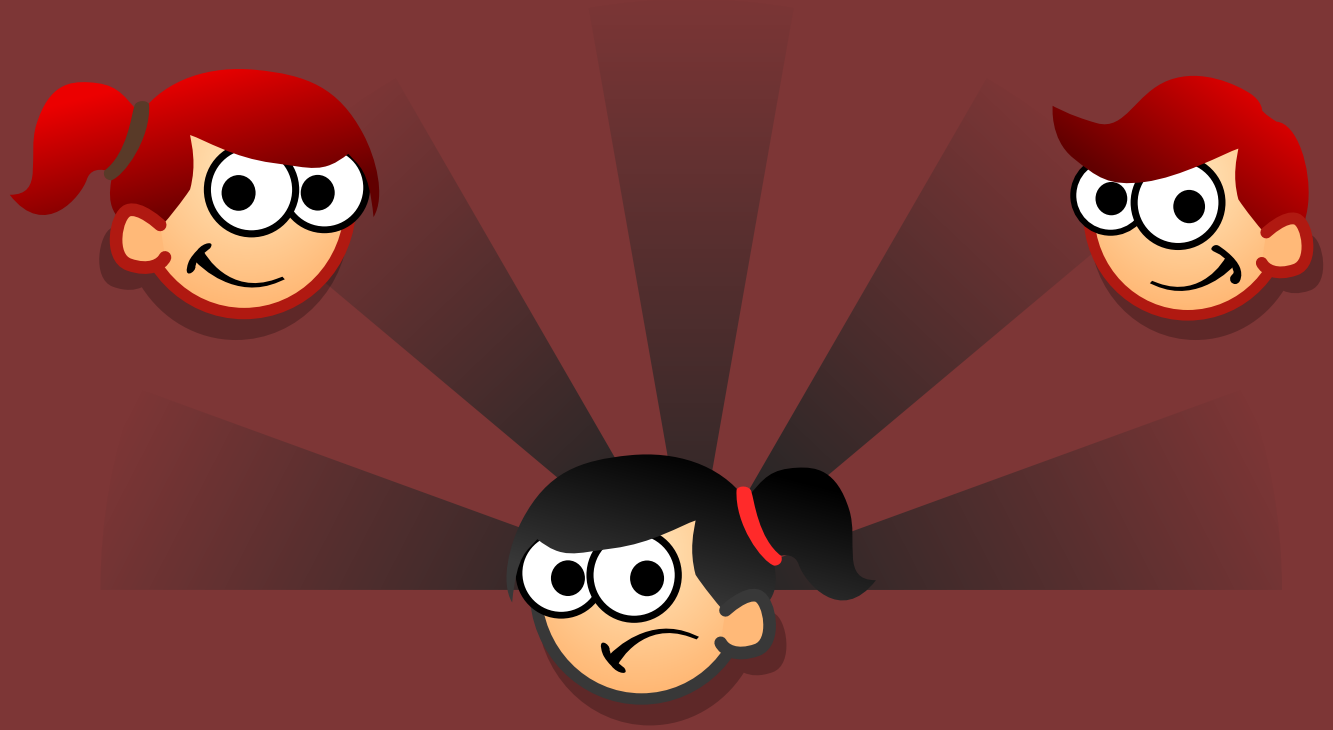


MViT-StrokeGAN Style Transfer

Group 3: Xingchu Zhang, Yiming Xu, Jiapeng He

Table of Contents

1. Problem Statement

What is the Problem We Are Trying to Solve?

Artistic Style Transfer:

- **Core Challenge**
  - Transform photographs into artistic styles
  - Preserve semantic content structure
  - Generate realistic artistic expressions
- **Specific Focus: Pencil Sketching**
  - Capture hand-drawn characteristics
  - Model directional stroke patterns
  - Maintain varied line weights
- **Technical Requirements**
  - Unpaired image-to-image translation
  - Real-time processing capability
  - High-quality artistic output

What is the Problem We Are Trying to Solve?

Problem Definition:

- Transform natural photos to pencil sketches
- No paired training data available
- Must preserve structural content
- Capture artistic stroke patterns

Key Challenges:

- Directional stroke synthesis
- Structure preservation
- Computational efficiency
- Quality vs speed trade-off

2. Motivation

Why is This an Important Problem? - Applications

Practical Applications:

- **Digital Art & Design**
  - Professional illustration workflows
  - Creative content generation
  - Rapid prototyping for artists
- **Educational Impact**
  - Art education tools
  - Skill development assistance
  - Accessible artistic creation
- **Entertainment Industry**
  - Concept art generation
  - Storyboard creation
  - Animation preprocessing

Real-World Impact:

- Replace time-intensive manual sketching
- Democratize artistic creation tools
- Enable non-artists to create sketches
- Accelerate creative workflows

User Benefits:

- Instant artistic transformation
- Consistent style application
- No artistic skill required
- Batch processing capability

Market Significance

Market Drivers:

- $2.3B AI art market growth
- Mobile app ecosystem demand
- Social media content creation
- Professional design tool needs

Future Implications:

- Foundation for other artistic styles
- Scalable creative AI platforms
- Enhanced human-AI collaboration

3. Technical Gap

Style Overfitting

Style overfitting represents one of the most critical challenges in neural style transfer, where models excessively replicate reference styles while lacking creative expression.

- **Key Manifestations:**
  - Lack of creativity, monotonous output generation
  - Rigid stylization, inability to adapt flexibly to diverse content
  - Over-reliance on specific style patterns from training data
- **Mathematical Root Cause:**

$$L_{style} = \sum_{l} \alpha_l ||G_l(\phi(I_{style})) - G_l(\phi(I_{generated}))||_F^2$$

  Excessive α_l weights cause style loss dominance over content preservation
- **Core Challenge:** Difficulty balancing style replication with creative expression

Text-Image Misalignment & Layout Instability

Layout Instability:

- **Artifacts:**
  - Structural distortions, loss of spatial coherence
- **Technical Origin:**

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Attention mechanism defects lead to inaccurate spatial relationship modeling

Computational Efficiency

Low Computational Efficiency:

- **Limitations:** Real-time application constraints, deployment bottlenecks
- **Contributing Factors:**
  - Model complexity with insufficient optimization
  - Multi-scale processing requirements
  - Memory-intensive attention computations

Domain Adaptation & Generalization Limitations

Domain Adaptation Challenges:

- **Cross-Domain Performance:** Models trained on specific domains fail to generalize
- **Style Domain Gaps:** Difficulty transferring between artistic styles (e.g., oil painting watercolor)
- **Content Domain Shifts:** Poor performance when content differs from training distribution

Quantitative Evaluation Challenges:

- **Subjective Quality Metrics:** No universally accepted quality measures
- **Evaluation Framework:**
  - Content preservation: LPIPS, SSIM scores
  - Style similarity: Gram matrix distances
  - Perceptual quality: FID, IS metrics

4. Method Analysis

MViT-StrokeGAN Framework

- Built upon CycleGAN with U-Net generator backbone
- Key innovations:
  - Dual MobileViT blocks for global context modeling
  - Dual stroke-aware modules for directional stroke synthesis
  - Composite objective function with novel loss terms
- Goal: Generate high-quality, stroke-consistent pencil sketches from natural images

Dual MobileViT Blocks for Global Context

- **Why MobileViT?**
  - Captures long-range dependencies and global context
  - More efficient than standard Transformers
  - Enhances feature representation for stroke patterns
  - Lightweight architecture suitable for real-time and edge applications
- **Dual Embedding Design:**
  - First block in downsampling path (128×128 feature map)
  - Second block in bottleneck (1×1 latent features)
  - Enables multi-scale contextual understanding
- **Implementation:**
  - Channel adaptation with 1×1 conv for 3-channel input
  - Spatial upsampling with pixel shuffling for skip connections
  - Hierarchical adaptation: early stages fixed, higher stages fine-tuned

Dual Stroke-Aware Modules for Directional Synthesis

- **Why Stroke-Aware Module?**
  - Explicitly models pencil stroke patterns
  - Minimal computational overhead
  - Fixed directional filters capture stroke characteristics
- **Dual Insertion Design:**
  - First module at 128×128 (early stage)
  - Second module at 32×32 (mid-level)
  - Captures coarse trends and refines at finer scales
- **Implementation:**
  - Four fixed directional filters (horizontal, vertical, 45°, 135°)
  - Filters initialized with Sobel-like patterns, non-trainable
  - Output concatenated with encoder features + 1×1 conv adapter

Composite Objective for Stroke and Structure

- **Total Loss:**

$$\mathcal{L}_{total} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{cycle}\mathcal{L}_{cycle} + \lambda_{idt}\mathcal{L}_{idt}$$

$$+\lambda_{stroke}\mathcal{L}_{stroke} + \lambda_{edge}\mathcal{L}_{edge} + \lambda_{grad}\mathcal{L}_{grad}$$

- **Balancing coefficients:** Determined empirically
- **Key components:**
  - Stroke consistency loss
  - Edge-preserving loss
  - Gradient loss

Stroke Consistency Loss ($\mathcal{L}_{stroke}$)

- **Dual-path design:**

$$\mathcal{L}_{stroke} = \lambda_{struct} \cdot \mathcal{L}_{struct} + \lambda_{style} \cdot \mathcal{L}_{style}$$

- **Structure-preserving term:**

$$\mathcal{L}_{struct} = \mathbb{E}_{x \sim p_{data}} \left[\| S(x) - S(G(x)) \|^2 \right]$$

  - Ensures coherence between input photo and generated sketch

- **Style-aligning term:**

$$\mathcal{L}_{style} = \mathbb{E}_{x \sim p_{data}, y \sim p_{sketch}} \left[\| S(y) - S(G(x)) \|^2 \right]$$

  - Encourages adoption of stroke patterns from real sketches

- **Benefit:** Retains input layout while adopting authentic stroke styles

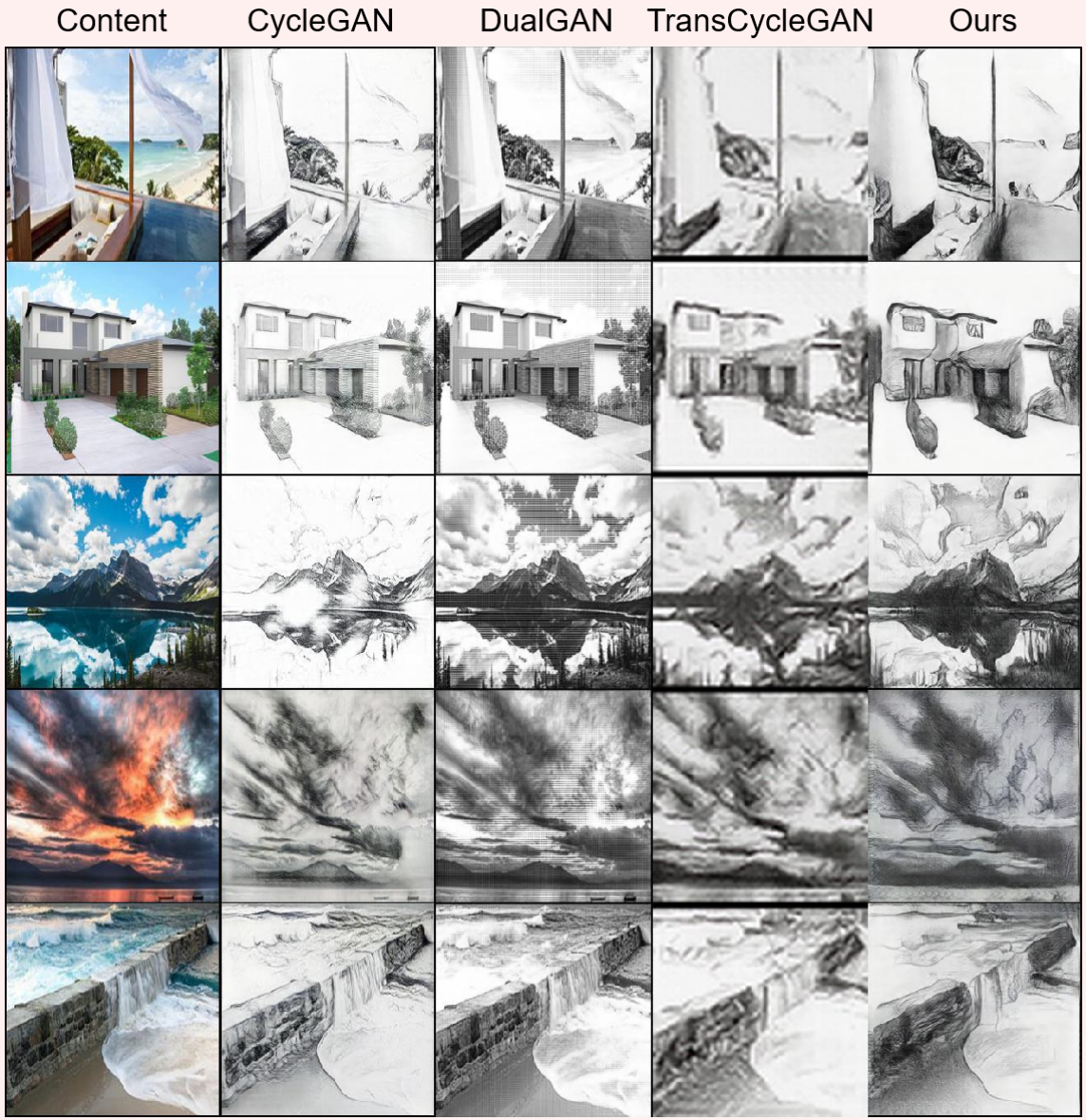
Ablation Study Setup

- **Model Variants:**
  - **Baseline+MViT:** U-Net + MobileViT only
  - **MViT-StrokeGAN w/o $\mathcal{L}_{grad}$:** Baseline+MVT + Stroke Module + $\mathcal{L}_{stroke}$ + $\mathcal{L}_{edge}$
  - **MViT-StrokeGAN (Ours):** Full model (w/o $\mathcal{L}_{grad}$ variant + $\mathcal{L}_{grad}$)
- **Evaluation Focus:**
  - Effectiveness of stroke-aware module and composite losses ($\mathcal{L}_{stroke}$, $\mathcal{L}_{edge}$)
  - Contribution of gradient loss ($\mathcal{L}_{grad}$) to detail preservation
  - Impact of weight ratios in stroke consistency loss ($\mathcal{L}_{stroke}$)

5. Qualitative Results

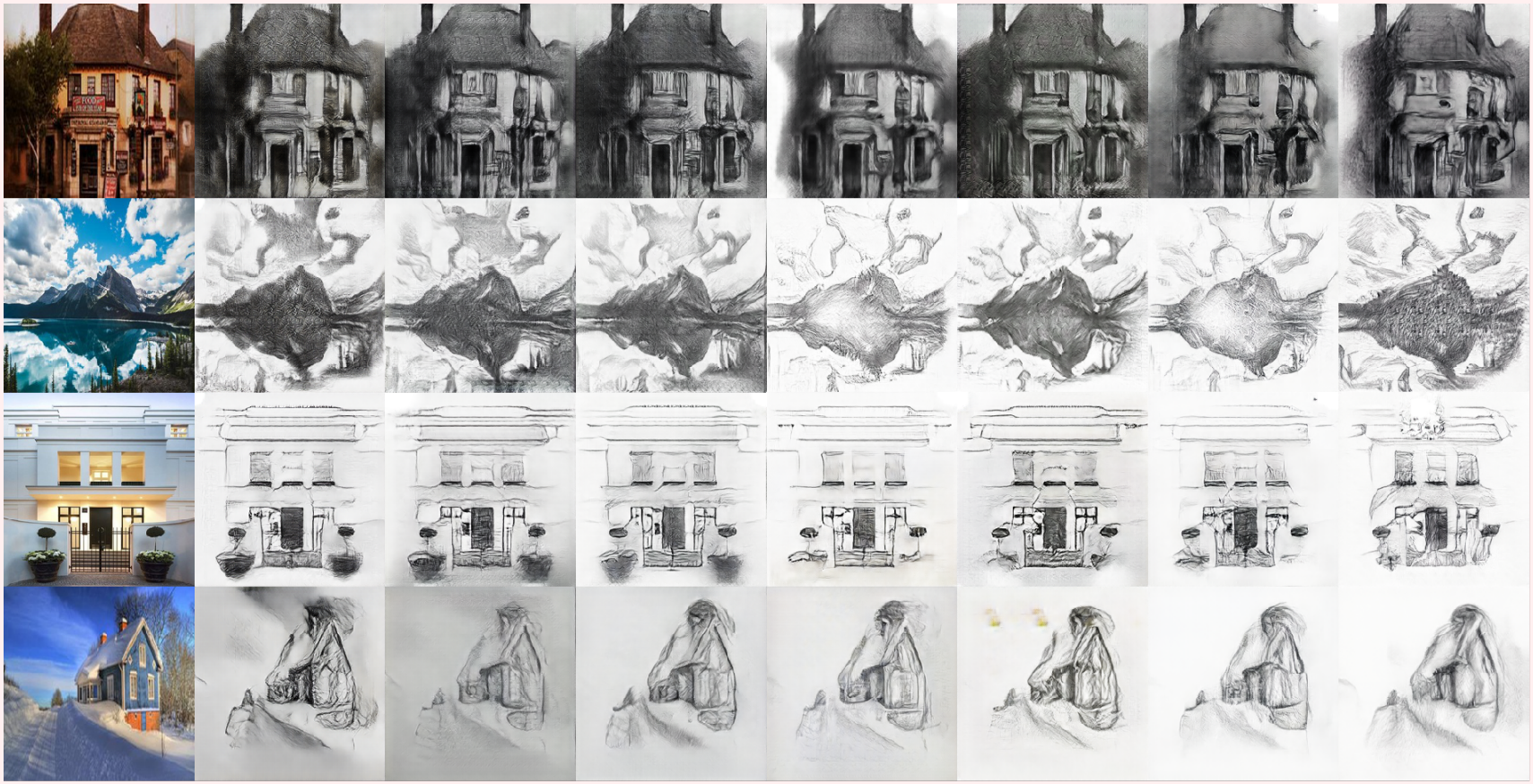
Ablation Study: Component Analysis

Qualitative Comparison with Baselines



- *Our method preserves input structure while generating realistic, coherent pencil strokes.*

Ablation Study: Stroke Loss Weight Ratio



1 : 0	3 : 1	3 : 2	1 : 1	2 : 3	1 : 3	0 : 1

Visual comparison of generated sketches using different weight ratios ($\lambda_{struct} : \lambda_{style}$) in the stroke consistency loss $\mathcal{L}_{stroke}$.

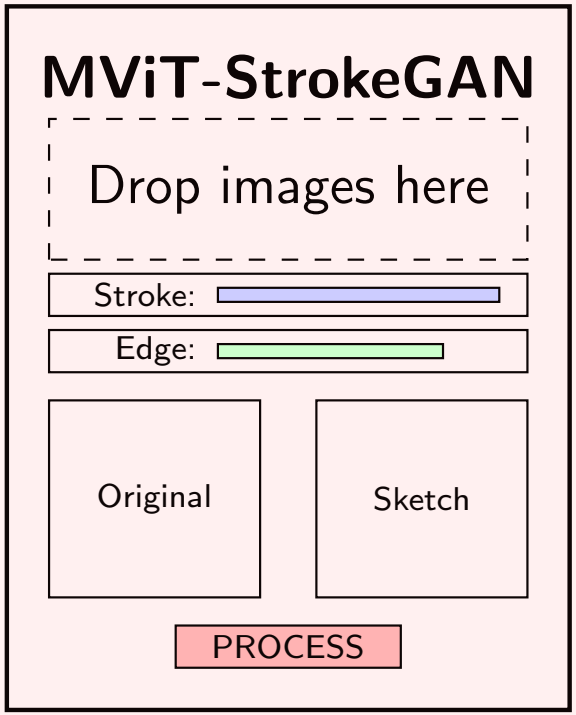
- The proposed MViT-StrokeGAN model significantly improves the quality of pencil sketch generation.
- Key components such as the stroke-aware module and gradient loss contribute to better structural fidelity and detail preservation.
- Balanced weight ratios in stroke consistency loss lead to optimal performance.

6. For end-users

End-User Interface

User-Friendly Features:

- **One-Click Processing**
  - Drag & drop image upload
  - Automatic preprocessing
  - Real-time preview
- **Quality Controls**
  - Stroke intensity slider
  - Edge sharpness adjustment
- **Batch Processing**
  - Multiple image handling
  - Consistent style application
  - Progress tracking



Deployment & Integration

Multi-Platform Deployment:

- **Web Application**
  - Browser-based interface
  - Cloud GPU processing
  - No local installation required
- **Mobile App**
  - iOS/Android compatibility
  - On-device processing option
  - Camera integration
- **API Service**
  - RESTful API endpoints
  - Third-party integration
  - Scalable cloud infrastructure

Target Users:

- Digital artists
- Architects
- Content creators
- Educational institutions
- ...

7. Conclusions

Key Achievements

Technical Innovations:

1. **MobileViT Integration**
   - First application in image-to-sketch translation
   - Lightweight transformer architecture
   - Multi-scale feature extraction
2. **Stroke-Aware Framework**
   - Novel stroke consistency loss
   - Content-adaptive texture generation
   - Directional pattern modeling
3. **Enhanced Loss Design**
   - Gradient consistency preservation
   - Edge-aware optimization
   - Multi-objective balance

Performance Highlights:

7× Faster than TransCycleGAN
Superior Edge Quality
Natural Stroke Patterns
Balanced Structure-Style

Future Directions

Short-term Extensions:

- **Multi-Style Support**
  - Watercolor paintings
  - Oil sketches
  - Charcoal drawings
- **Resolution Enhancement**
  - 512×512 and 1024×1024
  - Progressive training
  - Memory optimization
- **Interactive Controls**
  - Region-specific styling
  - User-guided refinement
  - Real-time adjustment

Long-term Vision:

- **Video Sketch Translation**
  - Temporal coherence
  - Motion-aware strokes
  - Real-time processing
- **Personalized Styles**
  - Artist-specific adaptation
  - Few-shot learning
  - Style interpolation

Contributions

Member's Name	**Common Contributions**	**Different Contributions**
Xingchu Zhang	• Determine the research topic and the model improvement plan. • Complete the writing of the paper. • Prepare for presentations.	Set baselines
Jiapeng He		Insert new modules
Yiming Xu		Set baselines

Thank You!

Questions?