

TL;DR

New methods aim to reduce identity drift across frames by fusing image, audio, video and text conditions.

Techniques include text-image fusion, hierarchical audio alignment and video-driven conditioning.

Specs and support vary by implementation; verify with official repos/papers before making promises.

1 The customization breakthrough that changes everything

May 8, 2025 saw community discussions of **HunyuanCustom** approaches to **subject consistency**. Results depend on datasets, prompts and hardware; verify with official sources.

1.1 The consistency challenge solved

| Problem | Traditional AI Video | Potential Approach |
|------------------------------|-----------------------------|--------------------------------------|
| Character drift | Face changes between frames | Temporal ID reinforcement |
| Multi-modal conflicts | Audio/visual misalignment | Hierarchical modality fusion |
| Style inconsistency | Random style variations | Reference-locked generation |
| Complex conditioning | Single input type only | 4-way multi-modal control |
| Memory requirements | 80GB+ VRAM needed | Optimizations, quantization (varies) |

2 Core architectural innovations

2.1 Text-Image Fusion Module (LLaVA-powered)

Unlike basic concatenation approaches, HunyuanCustom uses **LLaVA-based multi-modal understanding** for enhanced text-image alignment:

Multi-Modal Processing Pipeline:

- **Text Prompt** → LLaVA semantic embedding
- **Reference Image** → Visual feature extraction
- **Fusion Layer** → Cross-modal attention
- **Unified Representation** → Video generation

Why this matters: Traditional methods treat text and images as separate inputs, causing inconsistencies. LLaVA's joint understanding prevents modal conflicts.

2.2 Image ID Enhancement Module

The breakthrough **temporal concatenation technique** reinforces identity features across frame sequences:

ID Enhancement Workflow:

Identity Feature Extraction:

- **Input** → `extract_id_features(reference_image)`
- **Output** → `identity_features`

Frame-by-Frame Enhancement:

- **For Each Frame Index** → `range(len(video_frames))`
- **Temporal Concatenation** →
`concat_temporal_features(video_frames[frame_idx], identity_features,`
`temporal_weight=calculate_temporal_decay(frame_idx))`
- **Frame Update** → `video_frames[frame_idx] = enhanced_frame`

Return → `enhanced video_frames`

Result: Character faces remain **pixel-perfect consistent** even in complex motion sequences.

2.3 AudioNet Spatial Cross-Attention

For audio-conditioned generation, **AudioNet** achieves **hierarchical alignment** via spatial cross-attention mechanisms:

- **Low-level audio features** → Basic lip-sync and movement
- **Mid-level semantic content** → Emotion and expression mapping
- **High-level audio style** → Overall character animation consistency

- **Spatial cross-attention** → Frame-by-frame audio-visual alignment

2.4 Video-Driven Injection Network

The **patchify-based feature-alignment network** handles video conditioning by:

1. **Latent compression** of conditional video input
 2. **Patch-level feature extraction** for fine-grained control
 3. **Feature alignment** with target generation space
 4. **Injection** into the diffusion process at optimal layers
-

3 Multi-modal conditioning workflows

3.1 Image + Text conditioning

Use case: Brand mascot in different scenarios

Example Configuration:

Conditioning Inputs:

- **Image** → "brand_mascot.jpg"
- **Text** → "The friendly mascot waves hello in a sunny park setting"

Expected Output:

- **Design Consistency** → Maintains exact mascot design
- **Environment Application** → Applies park environment
- **Color Preservation** → Preserves brand color palette
- **Proportions** → Consistent character proportions

3.2 Audio + Video conditioning

Use case: Product demo with custom spokesperson

Multi-Modal Setup:

Condition Inputs:

- **Reference Video** → "spokesperson_sample.mp4" (3-second reference)
- **Audio Track** → "product_script.wav" (30-second narration)

- **Style Image** → "corporate_headshot.jpg" (Professional look)

Generation Parameters:

- **Duration** → 30 seconds
- **Resolution** → "720p"
- **Consistency Strength** → 0.95

Output → `result = hunyuan_custom.generate(conditions)`

References

- HunyuanVideo project: <https://github.com/Tencent-Hunyuan/HunyuanVideo>

3.3 All-modality conditioning

Advanced scenario: Interactive training video with multiple characters

Input Stack:

- **Character Images** → 3 people
- **Dialogue Audio Tracks** → Synchronized
- **Reference Video Style** → Corporate training
- **Text Descriptions** → Scene-by-scene
- **Emotion References** → Professional, friendly

Expected Output:

- **Duration** → 5-minute training video
- **Character Consistency** → Perfect identity preservation
- **Dialogue Quality** → Natural lip-sync
- **Visual Style** → Corporate appearance
- **Transitions** → Smooth scene changes

4 Production capabilities & performance

4.1 Technical specifications

| Feature | Specification | Impact |
|------------|----------------|-------------------------|
| Resolution | e.g., 1280x720 | Depends on model/config |

| | | |
|----------------------------|-----------------------------|-------------------------|
| Duration | e.g., ~30 seconds | Template dependent |
| GPU Memory | Implementation dependent | Verify per repo |
| Generation Time | Varies by hardware/settings | Trade speed/quality |
| Consistency Score | Varies by metric/dataset | Benchmark before claims |
| Multi-modal Support | Image/Audio/Video/Text | Where implemented |

4.2 Benchmark performance vs competitors

Professional evaluation across 500+ test scenarios:

| Metric | HunyuanCustom | Stable Video | Runway Gen-3 |
|--------------------------------|----------------------|-----------------|--------------------|
| ID consistency | Significantly higher | Baseline values | Improved stability |
| Text-video alignment | Strong alignment | Baseline values | Noticeable lift |
| Realism score | High realism | Baseline values | Clear upgrade |
| Multi-modal handling | ✅ Native | ❌ Limited | ⚠️ Basic |
| Custom subject fidelity | ✅ Excellent | ⚠️ Good | ⚠️ Good |

5 Real-world applications dominating

5.1 Brand content automation

Scenario: E-commerce brand with 500+ products

Traditional Workflow:

- **Model Hiring** → Separate budgets for each product category
- **Studio Shoots** → Multi-day productions with full crews
- **Post-Production** → Weeks of retouching and editing
- **Seasonal Reshoots** → Repeat investments each collection
- **Total Cost** → High six-figure spend and multi-month timelines

HunyuanCustom Workflow:

- **Reference Photos** → Brief portrait session with spokesperson
- **Script Templates** → One-day batch of prompts per category
- **Video Generation** → GPU time to batch 500 videos
- **Quality Review** → Short edit cycle to approve outputs
- **Total Cost** → Days instead of weeks, with spend concentrated on compute

ROI: Orders-of-magnitude reductions in cost and turnaround time when the pipeline is tuned correctly

5.2 Educational content scaling

Use case: Online course with consistent instructor across 100+ lessons

Before: Record all lessons in person = 3 months of instructor time

After: Record 10 reference lessons + generate remaining 90 = 1 week total

Consistency benefits:

- **Same instructor appearance** across all lessons
- **Consistent lighting and framing**
- **Professional audio quality** maintained
- **Easy content updates** without re-recording

5.3 Personalized marketing campaigns

Campaign: Insurance company with 50 regional representatives

Automated Regional Campaign Generation:

Setup:

- **Representatives Database** → `load_rep_database()` (50 people)
- **Campaign Script Template** → "Welcome to REGION insurance coverage..."

For Each Representative:

- **Image Input** → `rep.headshot`
- **Audio Generation** → `synthesize_voice(campaign_script, rep.voice_sample)`
- **Text Description** → "Professional insurance presentation for {rep.region}"
- **Style Reference** → `corporate_template`

Deployment:

- **Output** → `deploy_to_region(personalized_video, rep.region)`

Results: 50 personalized videos in 4 hours vs 2 months of individual recordings

6 Advanced customization techniques

6.1 Identity reinforcement strategies

Strong consistency (brand mascots, spokescharacters):

Strong Consistency Configuration:

- **Temporal Weight** = 0.95
- **Feature Injection Layers** = 2, 4, 6, 8
- **Consistency Loss Multiplier** = 2.0

Natural variation (human characters, realistic scenarios):

Natural Variation Configuration:

- **Temporal Weight** = 0.85
- **Feature Injection Layers** = 3, 6
- **Consistency Loss Multiplier** = 1.2

6.2 Multi-character scene management

Challenge: Maintaining multiple character identities simultaneously

Advanced Multi-Character Setup:

Host Character:

- **ID** → "host"
- **Reference Image** → "tv_host.jpg"
- **Audio Track** → "host_dialogue.wav"
- **Consistency Priority** → "high"

Guest Character:

- **ID** → "guest"
- **Reference Image** → "expert_guest.jpg"
- **Audio Track** → "guest_responses.wav"
- **Consistency Priority** → "high"

Scene Configuration:

- **Description** → "Professional interview setup with corporate backdrop"

- **Interaction Style** → "conversational"

Generation:

- **Output** → `interview_video = hunyuan_custom.generate_scene(scene_config)`
-

7 Production optimization & deployment

7.1 Hardware scaling recommendations

| Use Case | GPU Setup | Batch Size | Cost/Video |
|---------------------|------------------|--------------|---------------------|
| Development/Testing | RTX 4090 (24GB) | 1 video | GPU runtime |
| Small Business | RTX A6000 (48GB) | 2-3 videos | GPU runtime |
| Agency Production | A100 (80GB) | 4-6 videos | GPU + ops |
| Enterprise Scale | 4x A100 cluster | 12-16 videos | GPU + orchestration |

7.2 Quality optimization workflow

Production Quality Pipeline:

Phase 1 - Quick Preview Generation:

- **Quality** → "preview"
- **Duration** → 5 seconds
- **Resolution** → "480p"
- **Output** → `preview = hunyuan_custom.generate(config)`

Phase 2 - Client Approval Workflow:

- **Condition** → `if client_approves(preview)`

Phase 3 - Full Quality Generation:

- **Quality** → "production"
- **Duration** → 30 seconds
- **Resolution** → "720p"
- **Consistency Strength** → 0.95
- **Return** → `final_video` OR `request_revisions(preview)`

7.3 Content pipeline automation

Automated brand content factory:

Content Factory Pipeline:

Input Sources:

- **Brand Assets** → brand_assets/spokespersons/
- **Audio Scripts** → audio_scripts/product_categories/
- **Style References** → style_references/seasonal_campaigns/

Processing Rules:

- **Spokesperson Matching** → match_spokesperson_to_product_category
- **Style Updates** → apply_seasonal_style_updates
- **Multi-Resolution** → generate_multi_resolution_outputs

Output Destinations:

- **Social Media** → social_media/instagram_reels/
 - **Website** → website/product_pages/
 - **Email Campaigns** → email_campaigns/video_headers/
-

8 Integration with existing workflows

8.1 CMS & marketing automation platforms

WordPress/Drupal integration:

WordPress/Drupal Plugin Integration:

Function: generate_product_video(\$product_id)

Data Retrieval:

- **Product Data** → \$product = get_product(\$product_id)
- **Spokesperson** → \$spokesperson = get_brand_spokesperson()

Video Configuration:

- **Image** → \$spokesperson['headshot']
- **Text** → generate_product_script(\$product)
- **Audio** → synthesize_product_narration(\$product)

- **Style** → `get_brand_style_template()`

API Call:

- **Return** → `hunyuan_custom_api_call($video_config)`

Shopify app integration:

- **Auto-generate** product videos when new items are added
- **Batch update** existing products with video content
- **A/B test** different spokesperson/style combinations
- **Performance tracking** with conversion analytics

8.2 Video editing suite plugins

Adobe Premiere Pro extension:

- Import HunyuanCustom directly into timeline
- Real-time preview with different conditioning inputs
- Batch processing for multi-video projects
- Color correction presets for consistency

Final Cut Pro workflow:

- Custom effects library for HunyuanCustom integration
- Template projects with placeholders for quick generation
- Multi-cam editing for multi-character scenarios

9 Cost analysis & ROI calculations

9.1 Enterprise cost comparison

Scenario: Technology company creating 200 product demo videos annually

| Approach | Setup Cost | Per-Video Cost | Annual Total |
|-------------------------------|---------------|--------------------|----------------------------|
| Traditional Production | Highest | High | Highest |
| Stock Video + Editing | Moderate | Medium | Moderate |
| Synthesis/D-ID | Low (SaaS) | Usage-based | Lower |
| HunyuanCustom | Compute-first | Low once optimised | Lowest when fully utilised |

HunyuanCustom ROI:

- **Setup payback:** Track how many videos it takes to offset GPU + engineering costs.
- **Annual savings:** Calculate once you know your traditional spend; savings can be substantial.
- **Quality advantage:** Offers controllable visuals that can rival custom shoots when prompts and assets are dialed in.

9.2 Agency business model transformation

Before HunyuanCustom:

- Editors spend most of the week on manual timelines.
- Each project swallows multiple days.
- Throughput is capped by human hours.

After HunyuanCustom:

- Editors shift to quality control and final polish.
- Projects move from days to hours once templates are built.
- Weekly capacity expands because generation runs in parallel.

Business impact: Agencies report dramatic throughput gains without enlarging headcount when workflows are fully automated.

10 Advanced features & upcoming developments

10.1 Current capabilities (June 2025)

- ✓ **Audio-driven generation** via OmniV2V integration
- ✓ **Video-driven features** for style transfer
- ✓ **Single GPU support** (8GB VRAM minimum)
- ✓ **Batch processing** for production workflows
- ✓ **API endpoints** for programmatic access

10.2 Roadmap features

Q3 2025:

- **Real-time generation** for interactive applications

- **4K resolution support** with optimized models
- **Extended duration** (up to 2 minutes per generation)
- **Advanced emotion control** with micro-expression mapping

Q4 2025:

- **Multi-language consistency** across generated content
 - **Brand safety filters** for automated content screening
 - **Integration APIs** for major marketing platforms
 - **Mobile optimization** for on-device generation
-

11 Getting started: implementation guide

11.1 Technical setup (Week 1)

Installation and Setup:

Repository Setup:

- **Clone** → `git clone https://github.com/Tencent-Hunyuan/HunyuanCustom.git`
- **Navigate** → `cd HunyuanCustom`

Dependencies:

- **Install Requirements** → `pip install -r requirements.txt`
- **Install PyTorch** → `pip install torch torchvision --index-url https://download.pytorch.org/whl/cu118`

Model Weights:

- **Download** → `wget https://huggingface.co/tencent/HunyuanCustom/resolve/main/custom-model.safetensors`

Verification:

- **Test** → `python test_generation.py --config sample_config.yaml`

11.2 Content preparation (Week 2)

Asset organization:

Asset Organization Structure:

Characters Directory:

- **Spokesperson 01** → spokesperson_01.jpg
- **Spokesperson 02** → spokesperson_02.jpg
- **Brand Mascot** → brand_mascot.png

Audio Templates:

- **Product Intro** → product_intro_script.wav
- **Testimonial Template** → testimonial_template.wav
- **Call to Action** → call_to_action.wav

Style References:

- **Corporate Clean** → corporate_clean.mp4
- **Energetic Youth** → energetic_youth.mp4
- **Luxury Elegant** → luxury_elegant.mp4

Text Prompts:

- **Product Categories** → product_categories.json
- **Campaign Descriptions** → campaign_descriptions.json

11.3 Production workflow (Week 3-4)

Day-by-day implementation:

- **Week 3:** Single-video generation testing + quality optimization
- **Week 4:** Batch processing setup + team training
- **Month 2:** Full production integration + performance monitoring
- **Month 3:** Advanced features + custom fine-tuning

12 Community resources & support

12.1 Official resources

- **GitHub Repository:** [Tencent-Hunyuan/HunyuanCustom](#)
- **Model Hub:** [Hugging Face](#)
- **Research Paper:** [ArXiv](#)

- **Online Demo:** hunyuancustom.online

12.2 Professional services

Ready to implement HunyuanCustom for enterprise-scale customized video production? Our team specializes in **AI video infrastructure** for marketing and content teams.

Production teams:

DM us "**CUSTOM DEPLOY**" for a consultation on building your automated customized video pipeline with perfect subject consistency.

Last updated 25 Jul 2025. Model version: v1.0 (May 2025 release)