MUSE

Accepted at NeurIPS 2026 Main Track

NEURAL INFORMATION PROCESSING SYSTEMS 2026

Revisiting Diffusion Fine-Tuning for Unsupervised Domain Adaptation

One generator, multiple target domains

Xuan Qi · Yi Wei · Daniele Berardini · Vito Paolo Pastore · Vittorio Murino

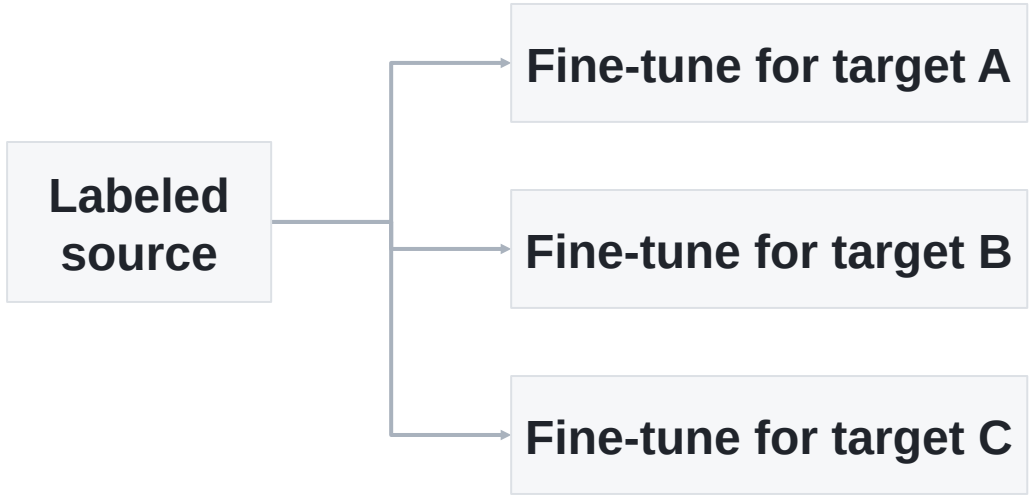
Istituto Italiano di Tecnologia · University of Genoa · Nanjing University

NEURAL INFORMATION PROCESSING SYSTEMS 2026

One labeled source, many unlabeled targets

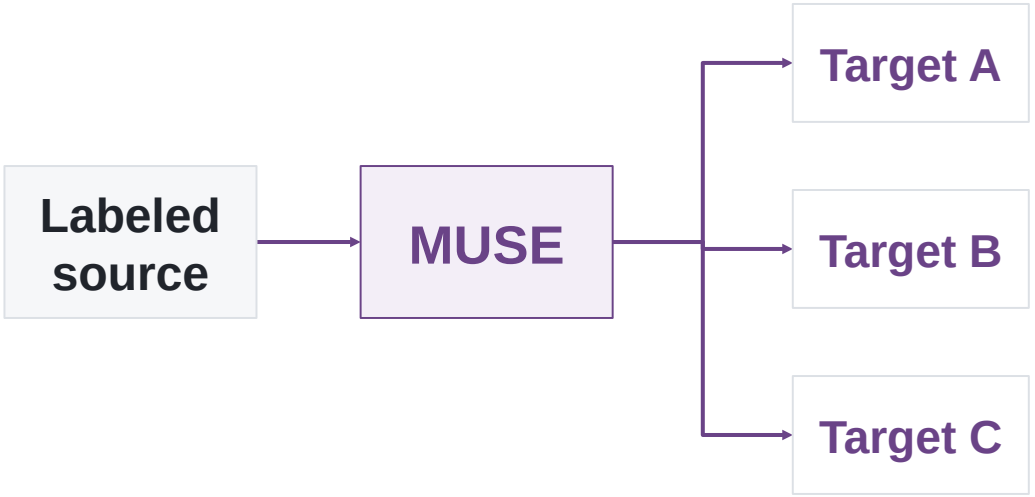
Conventional diffusion UDA repeatedly learns the same source semantics.

Pairwise fine-tuning



3 separate source–target adaptations

MUSE: shared adaptation

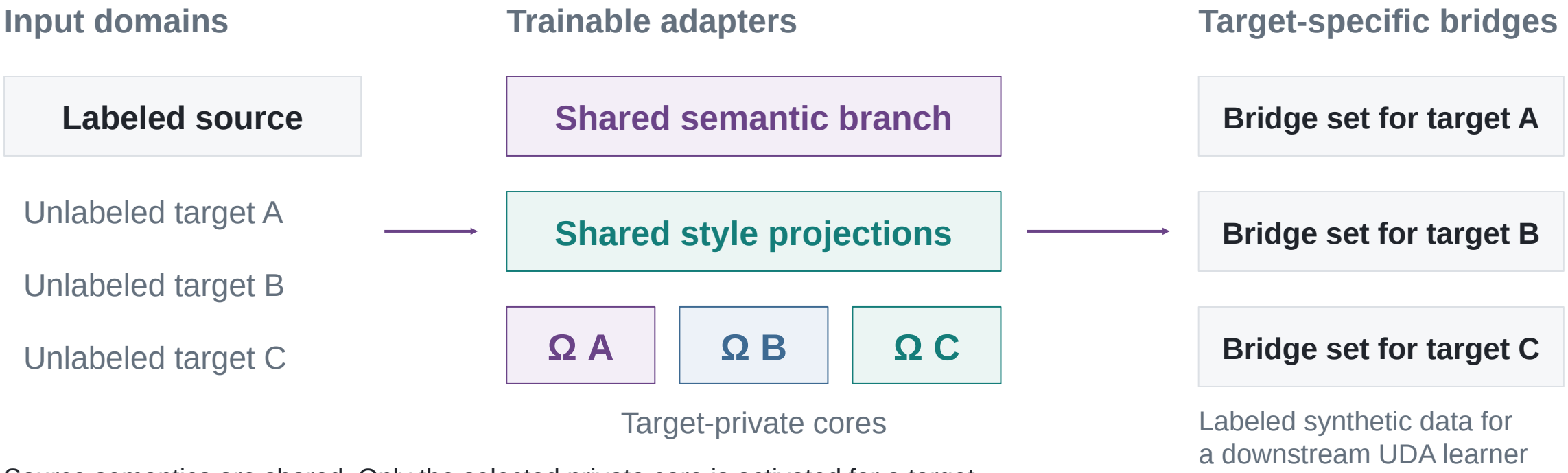


1 shared process + target-private cores

Can source-supervised semantic adaptation be learned once, then reused across targets?

Shared semantics + target-private style

Frozen SDXL backbone with lightweight adapters in U-Net attention projections.



Source semantics are shared. Only the selected private core is activated for a target.

During fine-tuning, target-distribution knowledge is shared across all target domains, providing richer target-domain knowledge than independent per-target fine-tuning.

NEURAL INFORMATION PROCESSING SYSTEMS 2026

Two passes, two distinct learning signals

Gradient routing prevents target updates from overwriting source semantics.

Pass 1: source

Labeled source image

Prompt: “a [class]”

Update shared semantic branch only

Stop-gradient on the semantic branch during the target pass

Pass 2: target

Unlabeled target image

Prompt: “an image”

Update style / private cores only

Orthogonality + magnitude regularization

Adaptive target sampling prioritizes harder target domains.

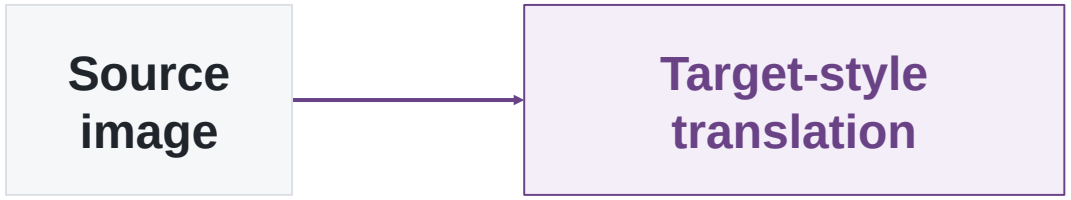
Two complementary ways to generate bridges

For each target, synthetic images carry labels into the downstream UDA stage.

Mode A

DDIM inversion / translation

Labeled source photo

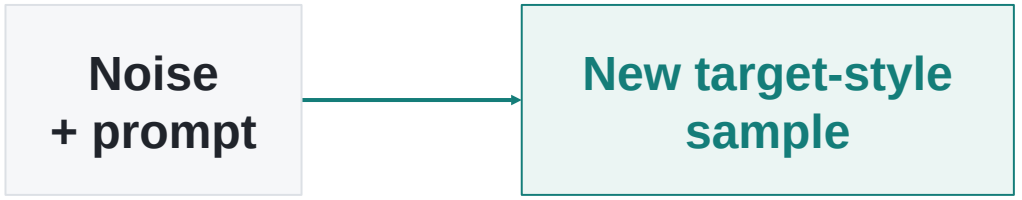


Source class label retained

Mode B

Class-conditional synthesis

Gaussian noise + class prompt



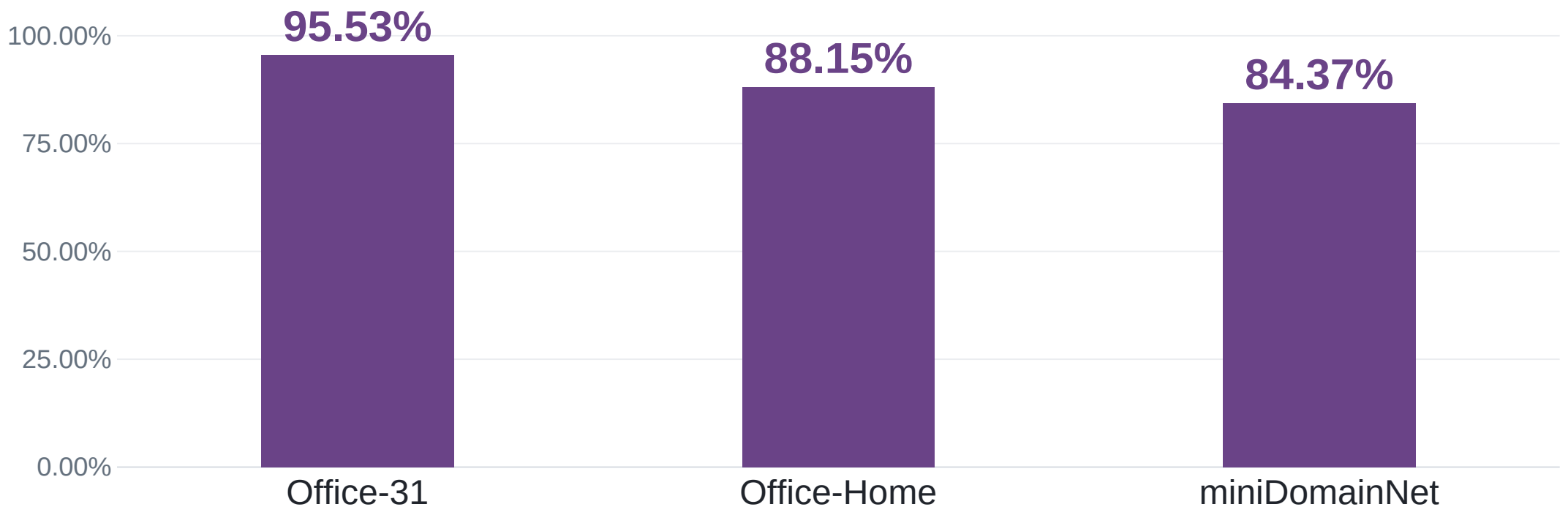
Class label from prompt

Target bridge set = translated samples + newly synthesized samples

NEURAL INFORMATION PROCESSING SYSTEMS 2026

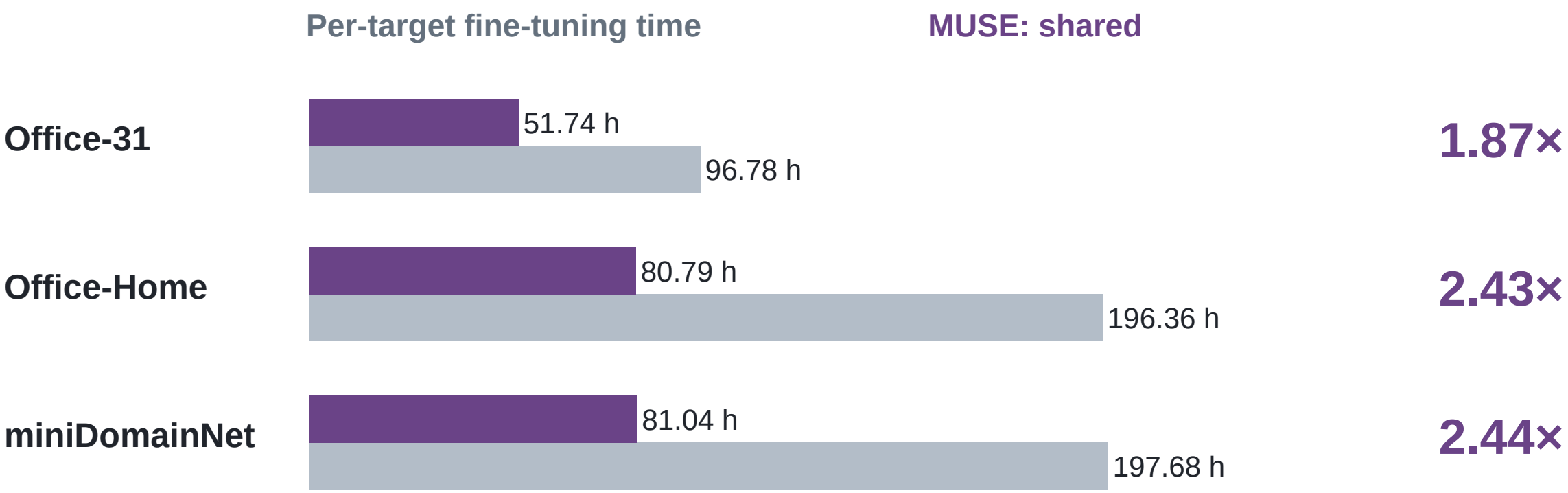
MUSE accuracy across three benchmarks

Target top-1 accuracy (%)



Up to 2.44× faster diffusion fine-tuning

Total generator adaptation time, SDXL on one A100 80 GB GPU.



One multi-target fine-tuning process replaces repeated per-target generator fine-tuning.

One generator, many targets

Reuse semantics.
Specialize style.
Improve adaptation.

MUSE improves the accuracy–efficiency trade-off for diffusion-based multi-target UDA.

Thank you!

Project page

xuanqi99.github.io/MUSE/