

Image-derived digital similarity matrices as a scalable proxy for pedigree and genomic relationship matrices

Masum Billah¹, M. Bermann¹, C. Y. Chen², B. Valente², E. Psota²
J. Holl², S. M. Bhandarkar³, I. Misztal¹, **D. Lourenco**¹

¹ Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA

² Genus PIC, Hendersonville, TN, USA

³ School of Computing, University of Georgia, Athens, GA, USA



College of Agricultural &
Environmental Sciences
UNIVERSITY OF GEORGIA



Research Objective and Motivation

- **Objective**

- Develop an image-based ML framework to estimate relationships among individuals:
 - **Genomic relationships (G)**
 - **Pedigree relationships (A)**

$$y = X\beta + Z\mathbf{u} + e$$
$$\begin{bmatrix} X' & y \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix}$$

- **Motivation**

- Missing or inaccurate pedigree records
- Limited availability of genomic data
- High genotyping costs
- Challenges in large-scale field data collection
- Data loss and record-management issues

- **Goal: Images as a proxy for genetic relatedness**

Overview of the Dataset

Large-scale dataset combining pedigree, image, trait, and metadata



- **Video Data**

- Controlled 2-D side-view videos
- 24K animals recorded at 60 fps
- Average duration: 7.19 s

- **Pedigree Records**

- Parent-offspring relationships
- Multi-generation pedigree

- **Genotype Data**

- 17.7K genotyped animals
- 50K SNP chip

- **Growth Traits**

- Body weight (BW)
- Loin depth (LD)
- Backfat thickness (BF)

- **Metadata**

- Contemporary group (CG)
- Birth year–month
- Farm, sex, litter

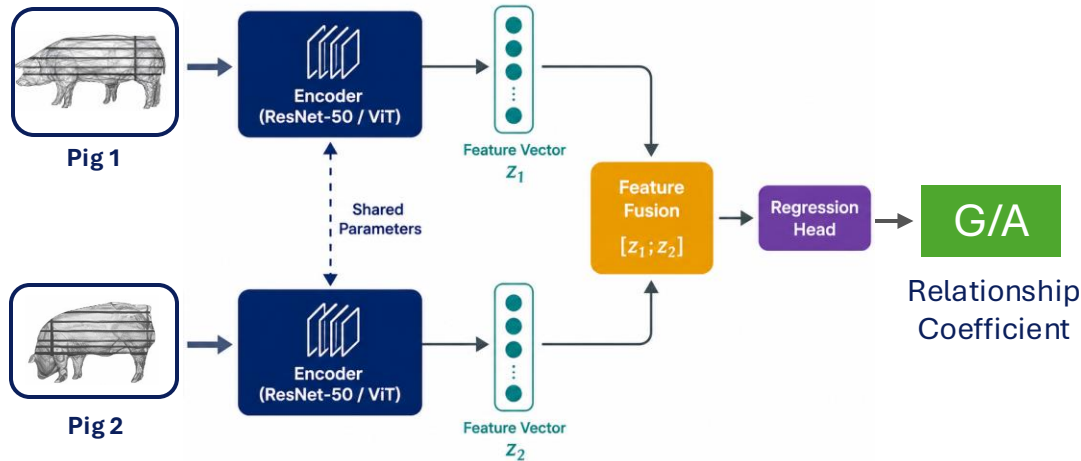
Key Statistics

Metric	Value
Animals	24K
Images Extracted	561.7K
Pedigree Records	15.9M
Genotyped Animals	17.7K
Collection Period	2021-2024

Modeling Approach 1: Supervised Learning

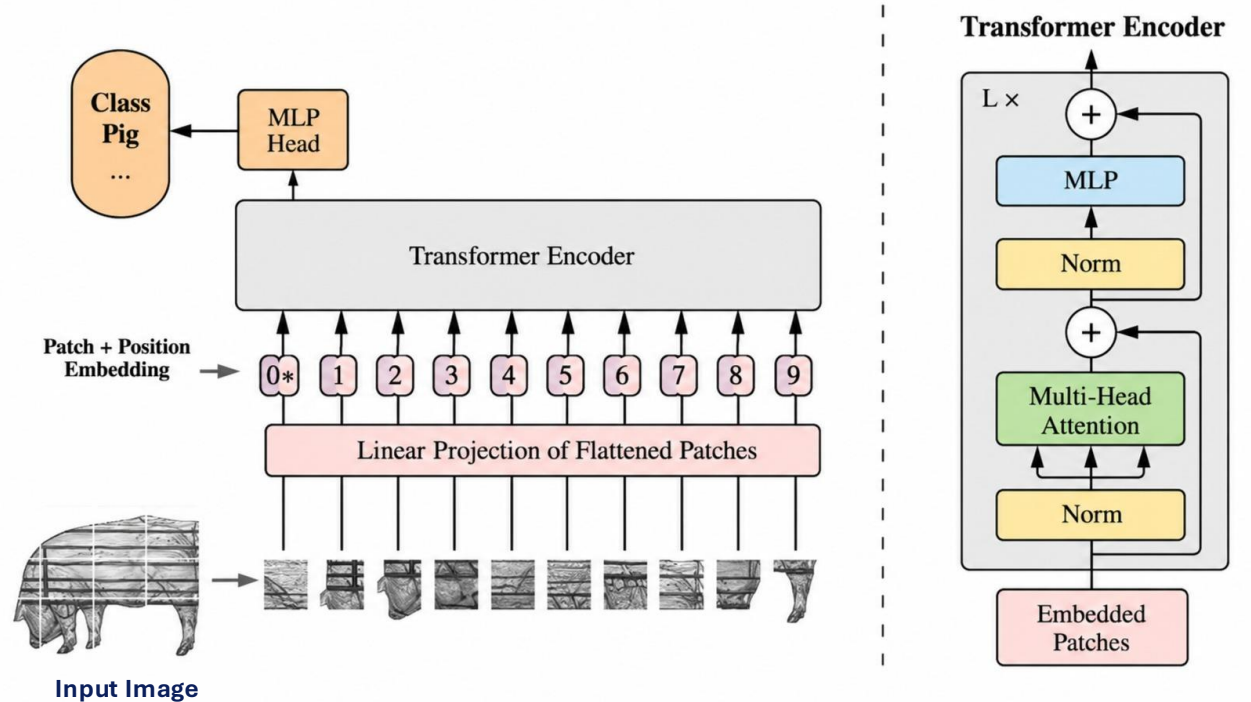
Learning genetic relationship coefficients from labeled image pairs

Siamese Training Setup



- Predict relationship coefficients from image pairs using shared-weight encoders and MSE loss
- Feature vectors are fused to capture pairwise similarity

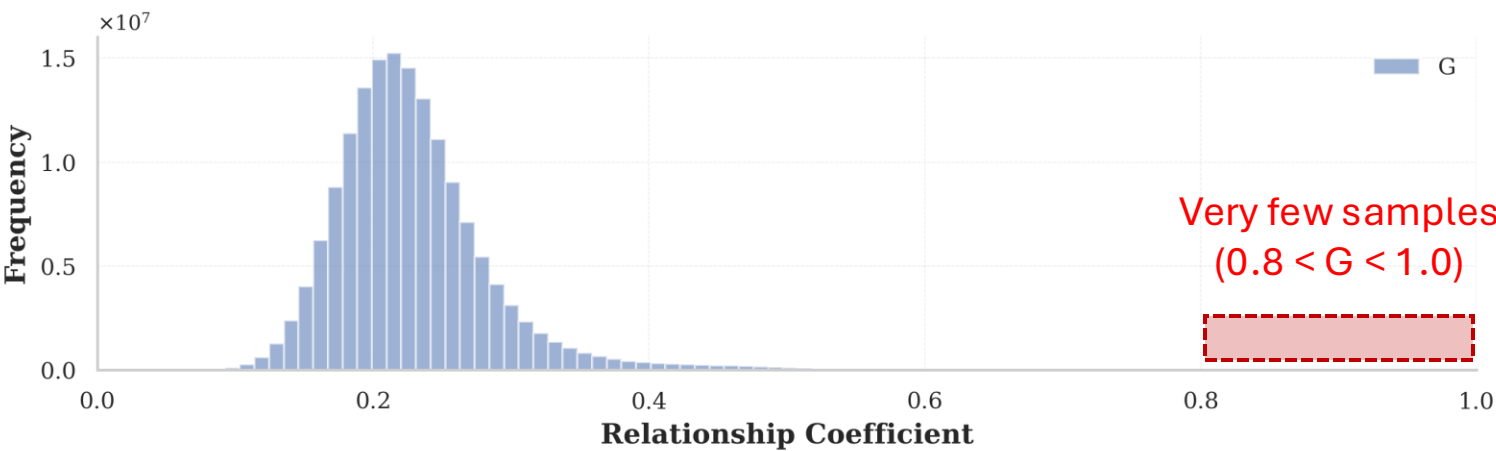
Vision Transformer Architecture



Target Distribution Imbalance

Most pairs are concentrated at low relationship coefficients

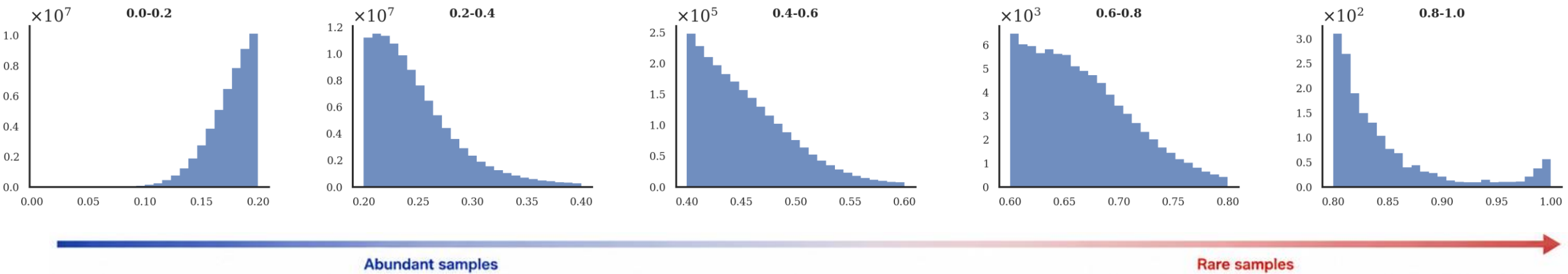
Overall Distribution of Genetic Relationship Coefficients



Key Statistics

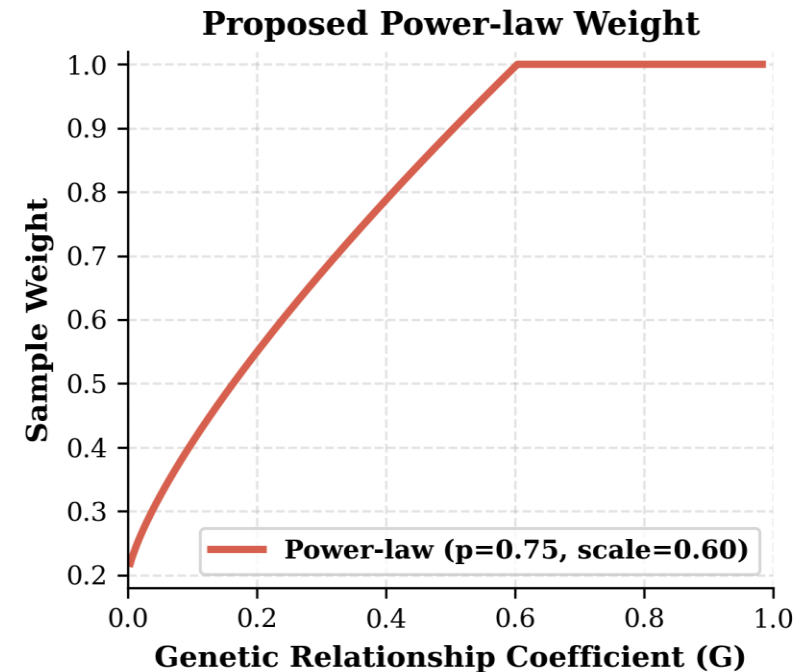
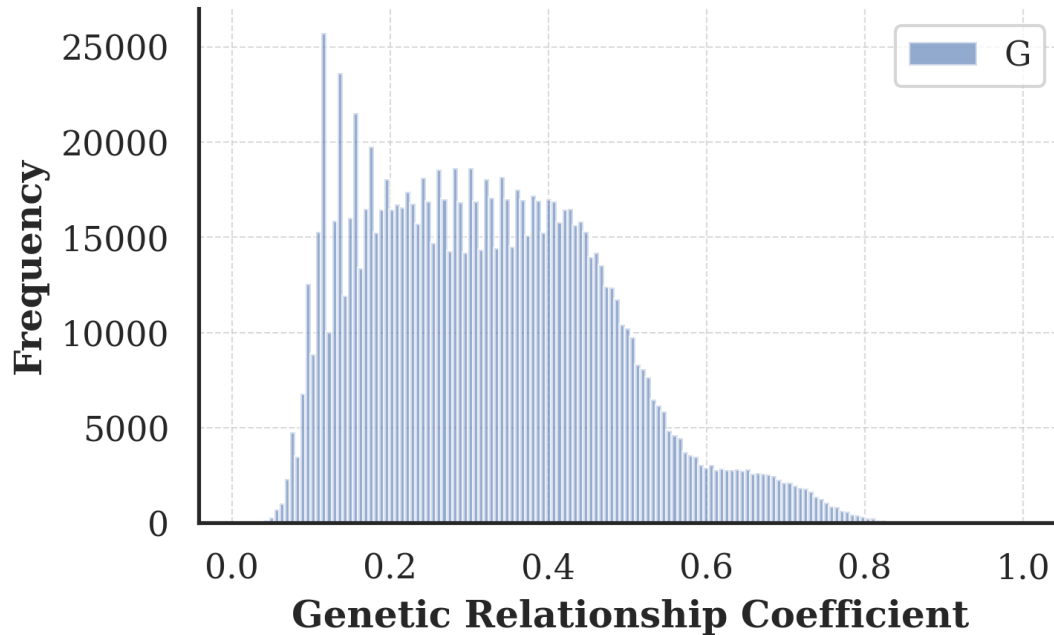
Metric	Value
Mean (G)	0.228
Median (G)	0.221
Majority Region	0.20–0.30
Rare Region	0.80–1.00
Distribution Type	Long-tailed
Imbalance ratio	> 7508:1

Decile-Wise Distribution Analysis



Distribution-Aware Sampling Strategy

Adaptive sampling and weighting were applied to mitigate target imbalance



Target distribution with adaptive sampling (left) and sample weight transformation (right)

- Maintains the overall target distribution while reducing severe imbalance
- Avoids over-representation of highly populated bins
- Preserves all samples from underrepresented relationship ranges
- Provides a balanced yet diverse training set

Evaluation Setup for Supervised Learning Models

Compared Configurations

Factor	Values
Backbones	Vit-S, ViT-B, ViT-L
Patch Sizes	16, 32
Fine-Tuning	U2, U4
Prediction Targets	G, A
Total Configurations	14

Vision Transformer Variants

Model	Patch Size	Embed Dim	Layers	Params
Vit-S/16	16	384	6	~22M
ViT-S/32	32	384	6	~22M
ViT-B/16	16	768	12	~86M
ViT-B/32	32	768	12	~86M
ViT-L/16	16	1024	16	~304M
ViT-L/32	32	1024	16	~304M

- **U2:** Unfreeze last 2 blocks
- **U4:** Unfreeze last 4 blocks

Performance of Supervised Learning Models

Table 1. Prediction of Genomic Relationship Coefficient (G)

Model	MSE	r
ViT-S/16 (U2)	0.0211	0.522
ViT-S/32 (U2)	0.0205	0.524
ViT-B/16 (U2)	0.0161	0.552
ViT-B/16 (U4)	0.0131	0.584
ViT-B/32 (U2)	0.0192	0.522
ViT-B/32 (U4)	0.0200	0.502
ViT-L/16 (U2)	0.0162	0.537

- ViT-Base/16 consistently achieved the strongest overall performance across both G and A22 prediction tasks
- Deeper fine-tuning (U4) improved genomic relationship prediction performance

Table 2. Prediction of Pedigree Relationship Coefficient (A)

Model	MSE	r
ViT-S/16 (U2)	0.0169	0.543
ViT-S/32 (U2)	0.0164	0.509
ViT-B/16 (U2)	0.0116	0.602
ViT-B/16 (U4)	0.0118	0.588
ViT-B/32 (U2)	0.0141	0.546
ViT-B/32 (U4)	0.0138	0.596
ViT-L/16 (U2)	0.0111	0.570

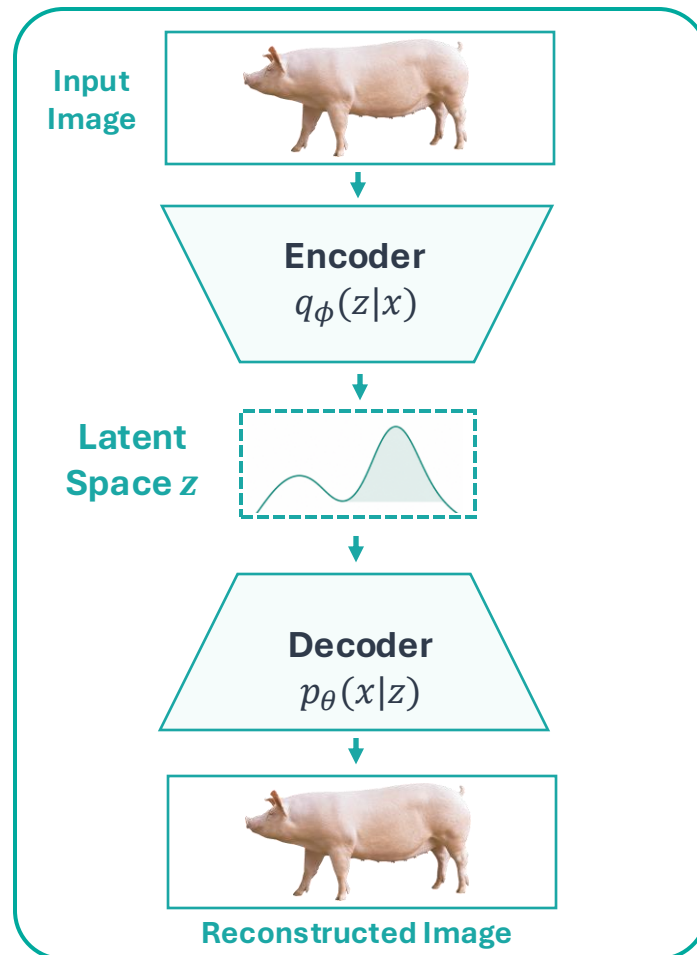
- Moderate correlations ($r \approx 0.58\text{--}0.60$) suggest side-view body images capture phenotypic features associated with genetic relationships

Modeling Approach 2: Self-Supervised Learning

Self-supervised representation learning from unlabeled images

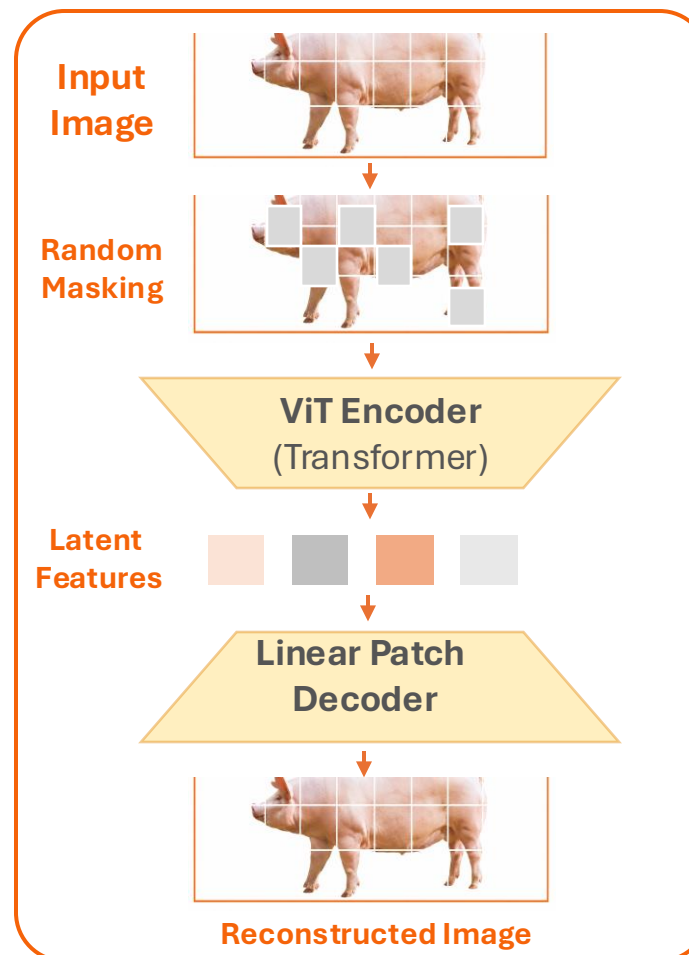
A) β -VAE

Reconstruction Learning



B) Masked ViT-AE

Masked Image Reconstruction



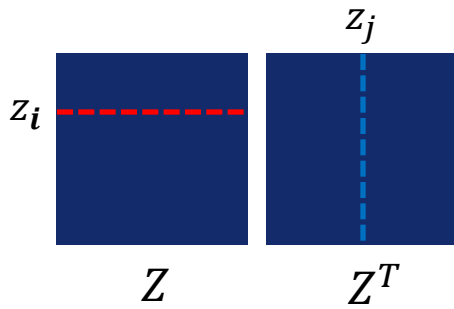
- Learns disentangled latent variables via KL-regularized reconstruction
- Image reconstruction from compressed embeddings using custom CNN model
- Applies random patch masking during self-supervised training
- Reconstruction of masked patches using context

Similarity Measures in Latent Space

Relationship coefficients are estimated from image-derived latent representations

A) Dot Product

$$D = \frac{1}{m} Z Z^T \quad \text{then} \quad D_{ij} = \frac{1}{m} z_i^T z_j$$

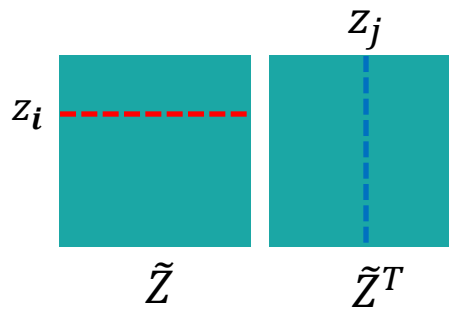


$$z_i \cdot z_j = \| z_i \| \| z_j \| \cos(\theta)$$

- Z contains raw latent features
- Captures magnitude and direction
- Sensitive to vector scale

B) Cosine Similarity

$$D = \tilde{Z} \tilde{Z}^T \quad \text{then} \quad D_{ij} = \frac{z_i^T z_j}{\| z_i \| \| z_j \|}$$

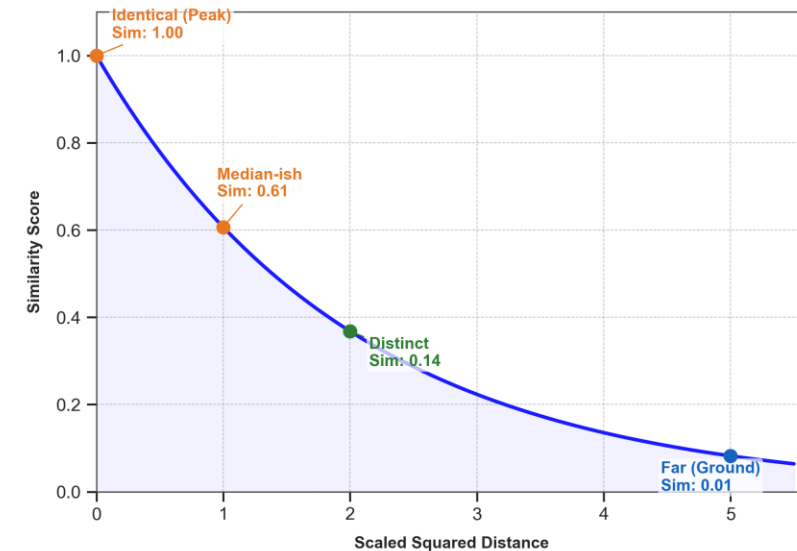


$$\frac{z_i \cdot z_j}{\| z_i \| \| z_j \|} = \frac{\| z_i \| \| z_j \| \cos(\theta)}{\| z_i \| \| z_j \|} = \cos(\theta)$$

- Latent features normalized to unit length
- Measures angular similarity
- Scale invariant

C) RBF Similarity

$$D_{ij} = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right)$$



- Captures non-linear relationships
- Produces bounded similarity values

Relationship Prediction Performance

Performance of Linear, Cosine, and RBF Similarity Kernels Across G and A Targets

Table 3. Accuracy of image-derived similarity matrices. Comparing three kernel formulations against pedigree (Ped.) and genomic (Geno.) benchmarks.

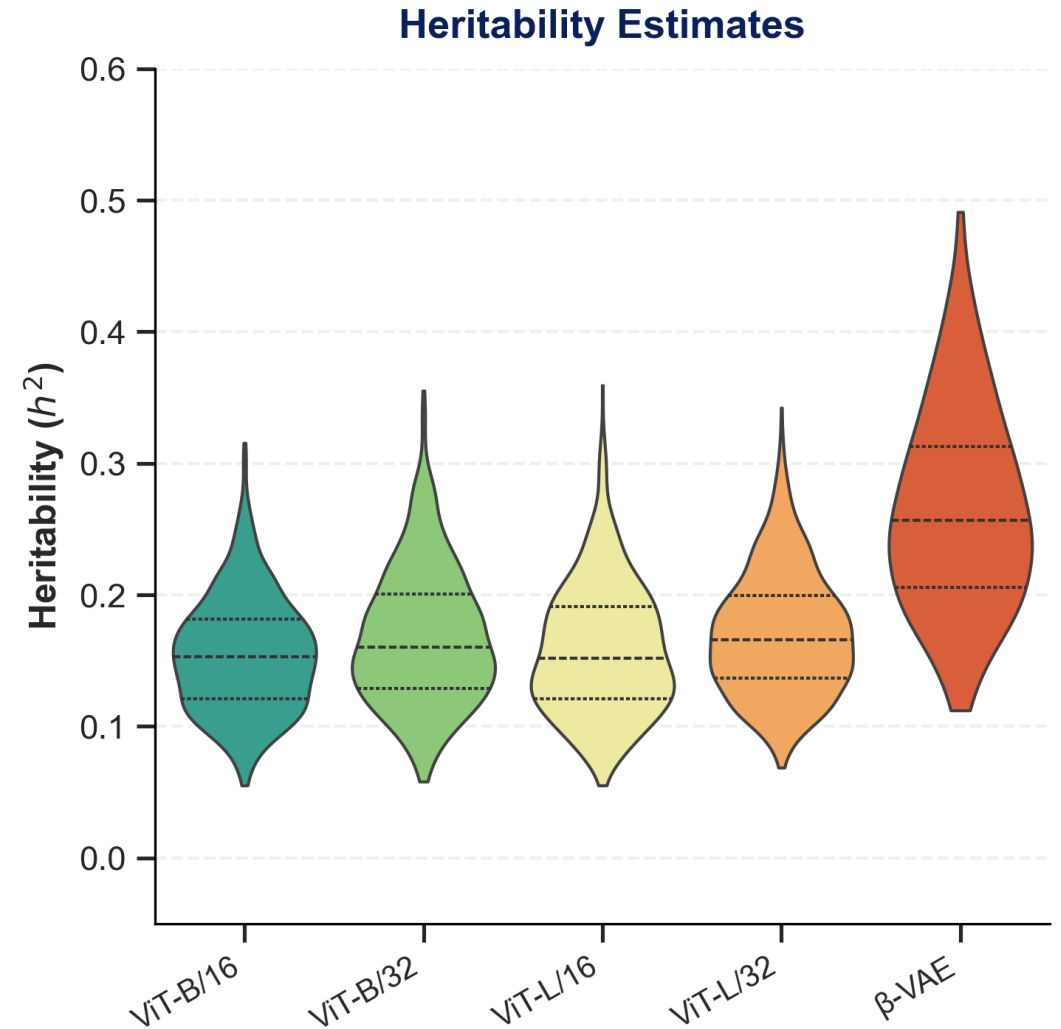
Model Architecture	D_{linear}		D_{cos}		D_{rbf}	
	Ped.	Geno.	Ped.	Geno.	Ped.	Geno.
ViT-B/16	0.25	0.23	0.28	0.25	0.25	0.20
ViT-B/32	0.27	0.25	0.31	0.27	0.29	0.22
ViT-L/16	0.24	0.22	0.26	0.25	0.23	0.20
ViT-L/32	0.25	0.23	0.30	0.26	0.27	0.20
β -VAE	0.11	0.12	0.11	0.12	0.06	0.08
MAE	0.09	0.10	0.06	0.10	-0.04	0.05

Genetic Relevance of Latent Representations

Heritable latent representations capture biologically meaningful genetic variation

Heritability Analysis

- **ViT latent features:** $h^2 = 0.06\text{--}0.36$ (mean = **0.17**)
- **β -VAE latent features:** $h^2 = 0.11\text{--}0.49$ (mean = **0.27**)
- β -VAE latent space exhibits higher heritability, suggesting stronger alignment with genetic variation



Comparison Summary: Supervised vs. Self-Supervised

Comparison of model performance for G and A targets

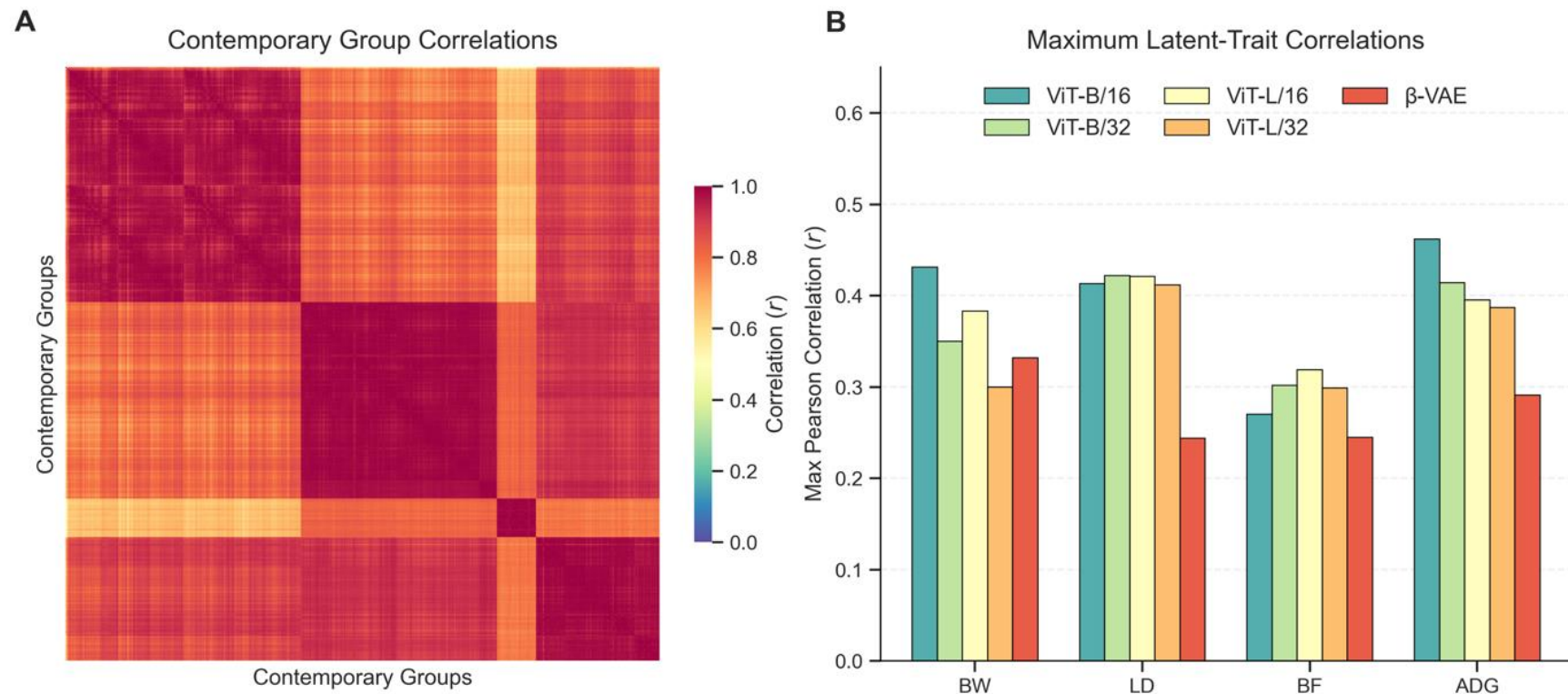
Model Architecture	Target	r
ViT-B/16 (U2)	A	0.602
β-VAE	A	0.11
Masked ViT-AE	A	0.09
ViT-B/16 (U4)	G	0.584
β-VAE	G	0.12
Masked ViT-AE	G	0.10

Table 4. Comparison of learning approaches for relationship prediction

- ViT embeddings achieved the highest correlation with both G and A targets
- Image-derived representations contain predictive information for relationship estimation
- β-VAE latent features retained modest but measurable biological signal

Attribute-Specific Correlation Analysis

Latent representations reveal distinct clustering patterns across animal attributes



- Distinct clustering among contemporary groups
- High within-group similarity indicates consistent latent representations
- Latent features moderately correlate with BW, LD, BF, and ADG
- Maximum latent-trait correlations ranged from 0.42–0.46

A) Contemporary Group (CG) Correlations. **B)** Maximum Latent–Trait Correlations

Conclusions

- Deep learning models extracted latent representations from 2D pig images that captured both genetic and environmental information
- β -VAE latent features exhibited higher heritability, suggesting stronger representation of underlying genetic variation
- ViT embeddings achieved the highest accuracy for predicting genomic (G) and pedigree (A) relationships, indicating greater ability to capture individual phenotypic differences
- Phenomic similarity derived from images was associated with both pedigree and genomic relationships
- Future work will incorporate image-derived similarity matrices into the genomic evaluation process

Acknowledgements

UGA Team



PIC Team

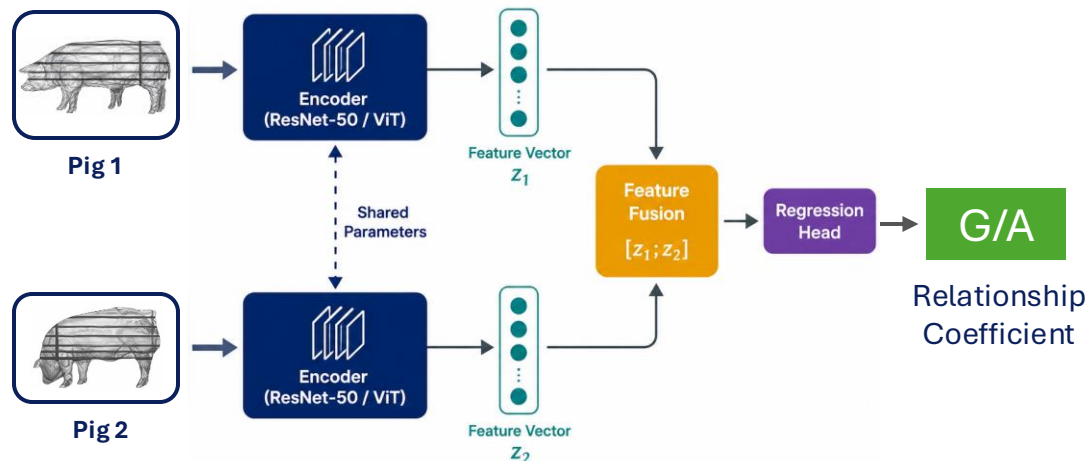


Thank You
Questions?

Overview of Modeling Approaches

- **Approach 1: Supervised Learning**

- Predict relationship coefficients from image pairs using shared-weight encoders and MSE loss
- Shared encoder extracts features from paired images
- Feature vectors are fused to capture pairwise similarity
- **Vision Transformers (ViT):** Learns global representations through self-attention



- **Approach 2: Self-Supervised Learning**

- Learn latent representations from unlabeled images
- Extract latent features using the trained encoder
- Construct latent-space similarity matrices for genomic evaluation
- **B-VAE:** Reconstruction-based representation learning
- **Masked ViT-AE :** Masked-patch reconstruction learning

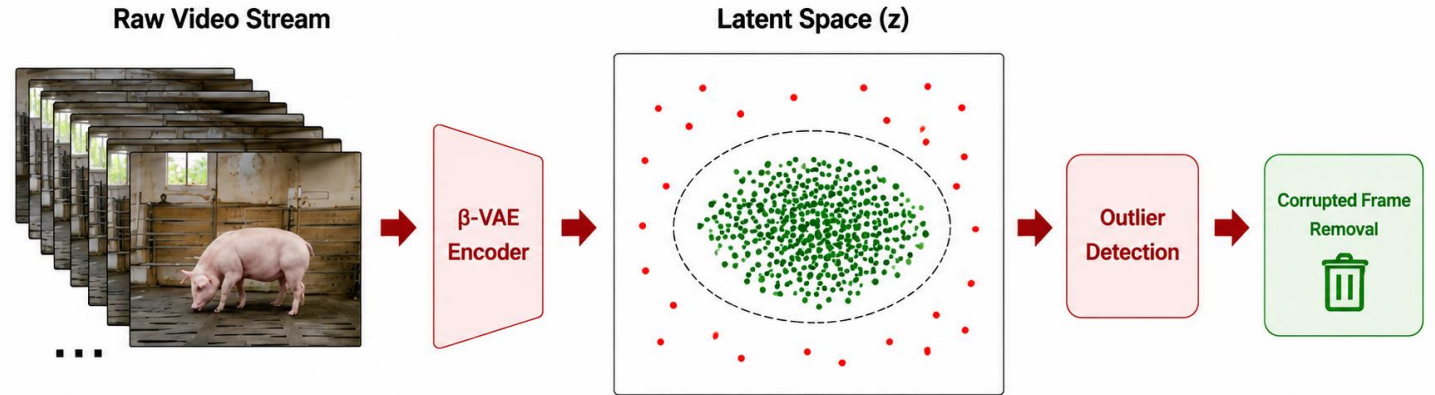


Quality Control of Image Data

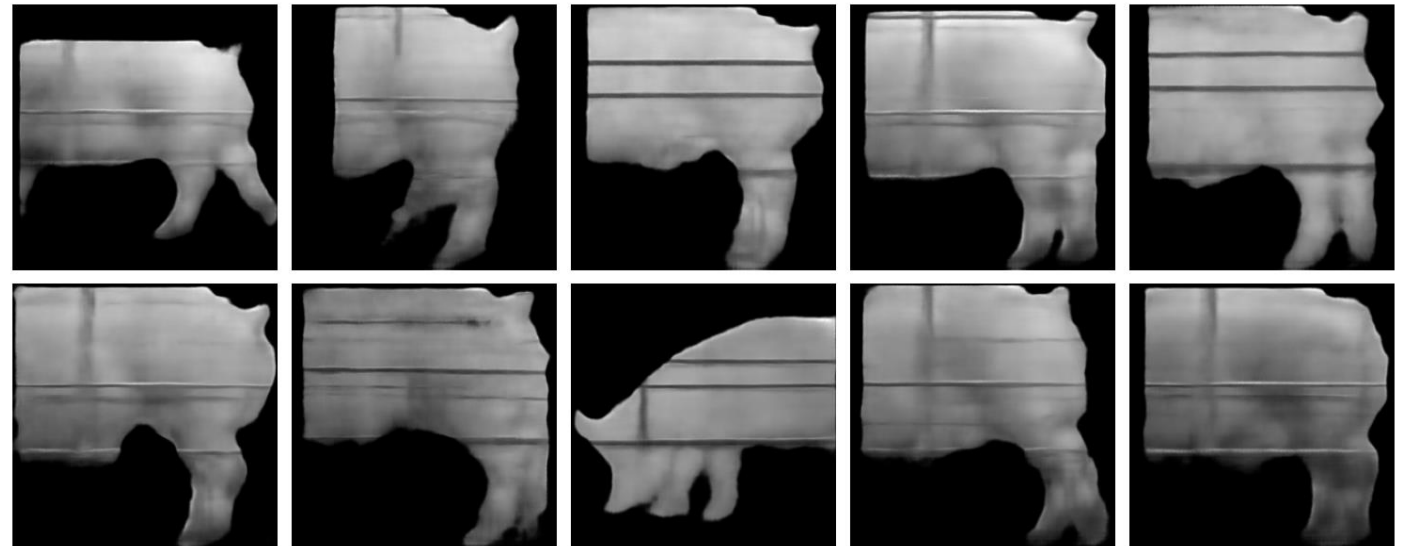
Detection and removal of low-quality frames using β -VAE

- Identified partial crops, occlusions, and corrupted frames
- Low reconstruction quality served as an indicator of image anomalies
- Detected corrupted frames were excluded from downstream analyses
- Improved reliability of relationship estimation

β -VAE-Based Quality Control Pipeline



Detected Corrupted Frames

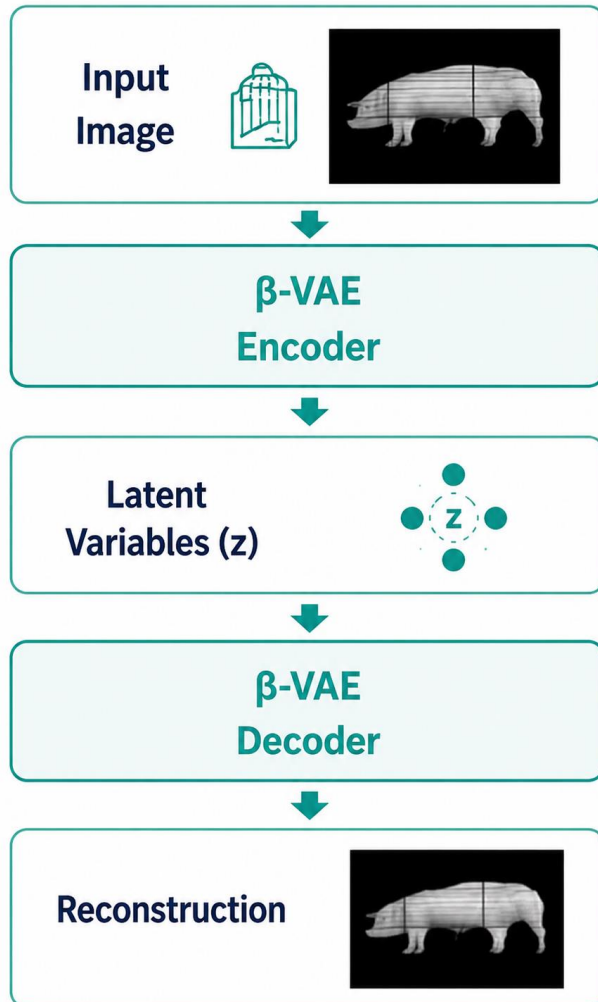


Key Result

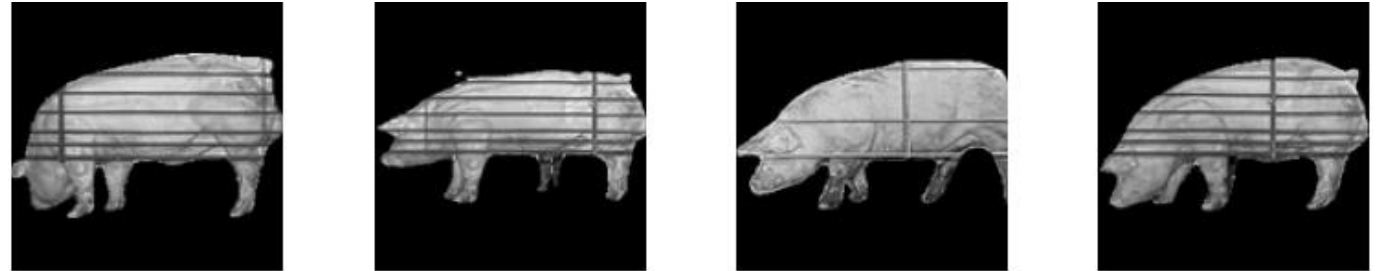
Removed **5.8%** of low-quality and corrupted frames.

β -VAE Reconstruction Quality

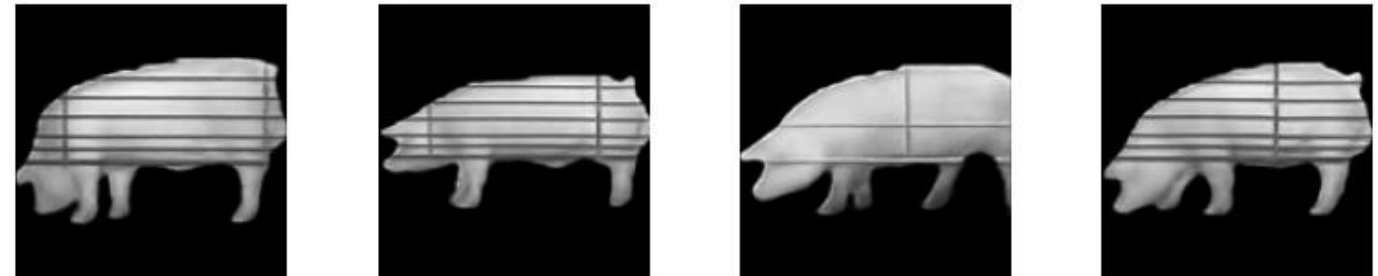
Latent variables preserve biologically relevant shape while suppressing noise



Original Images



β -VAE Reconstructions



β -VAE Reconstructions

- **Preserve** biologically relevant morphology
- **Capture** pose and structural information
- **Suppress** noise and background artifacts