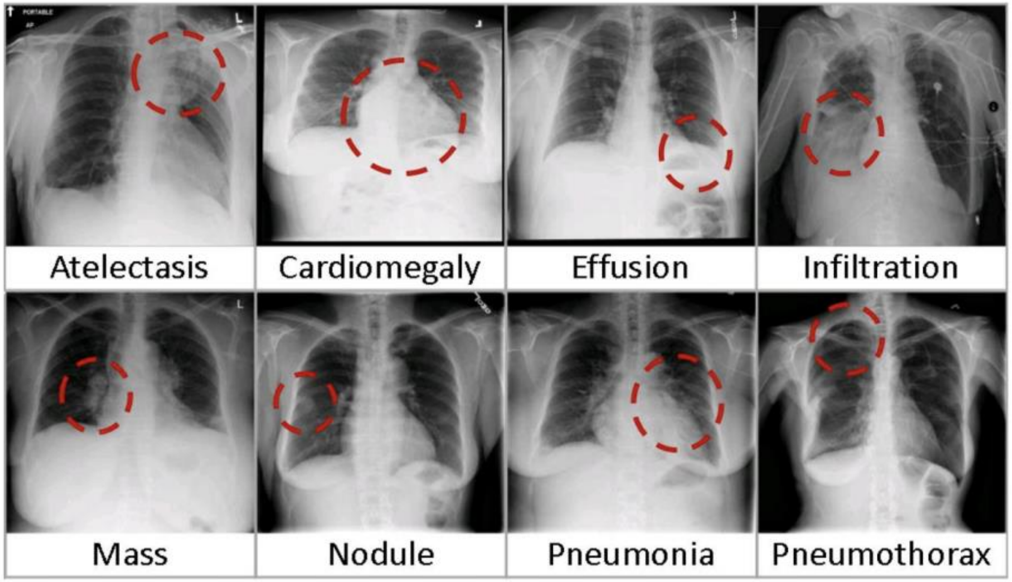
Medical Image Analysis: Discriminative and Generative Models

From Segmentation to Latent Representation



Suchen Zheng

CBB 752 Spring 2026

Why Biomedical Image Analysis Matters

3.6 Billion

Diagnostic imaging procedures performed globally each year (WHO, 2023)

>17,000

Projected shortage of radiologists in the U.S. by 2030 (ACR Report)

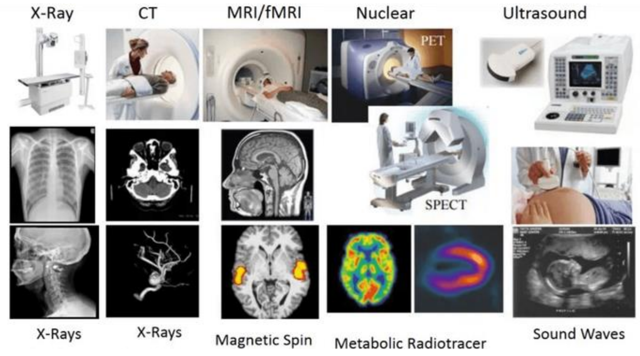
30%

Disagreement rate between expert readers on complex cases

- Manual image interpretation is time-consuming, subjective, and prone to inter-observer variability.
- AI offers faster, consistent, and scalable diagnostic support, especially for under-resourced settings.
- FDA has cleared **>950 AI-enabled algorithms** by mid-2025, making radiology the top AI application in healthcare.

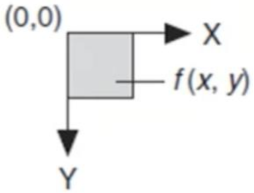
The Landscape of Biomedical Imaging Modalities

MODALITY	PHYSICAL PRINCIPLE	TYPICAL APPLICATIONS	KEY CHALLENGES FOR AI
X-Ray / CT	Ionizing radiation attenuation	Chest pathology, bone fractures, lung nodules	Low contrast for soft tissue, 3D reconstruction artifacts
MRI	Nuclear magnetic resonance	Brain tumors, cardiac function, musculoskeletal	Long acquisition times, multi-sequence heterogeneity
Ultrasound	High-frequency sound waves	Fetal imaging, cardiac echo, liver lesions	Speckle noise, operator dependency, shadowing
PET / SPECT	Radiotracer metabolic uptake	Oncology staging, neurodegeneration tracking	Low spatial resolution, anatomical co-registration
Histopathology	Light microscopy of stained tissue	Cancer grading, tumor microenvironment	Gigapixel whole-slide images (WSI), stain variability
Fundus / OCT	Optical imaging of the retina	Diabetic retinopathy, glaucoma, AMD	Domain shift across devices, image quality issues



Images as Numerical Tensors

- A digital image: $x \in \mathbb{R}^{H \times W \times C}$
- Properties:
  - Resolution: spatial dimensions
  - Bit depth: intensity precision
  - Channels: grayscale (1), RGB (3)
  - In medical imaging: often 3D tensors
    - H×W×D×C, D is depth in the z-direction
- Padding concept for convolution:
  - Zero padding (B)
  - copying of edge pixel values (C)
  - Mirror padding (D)



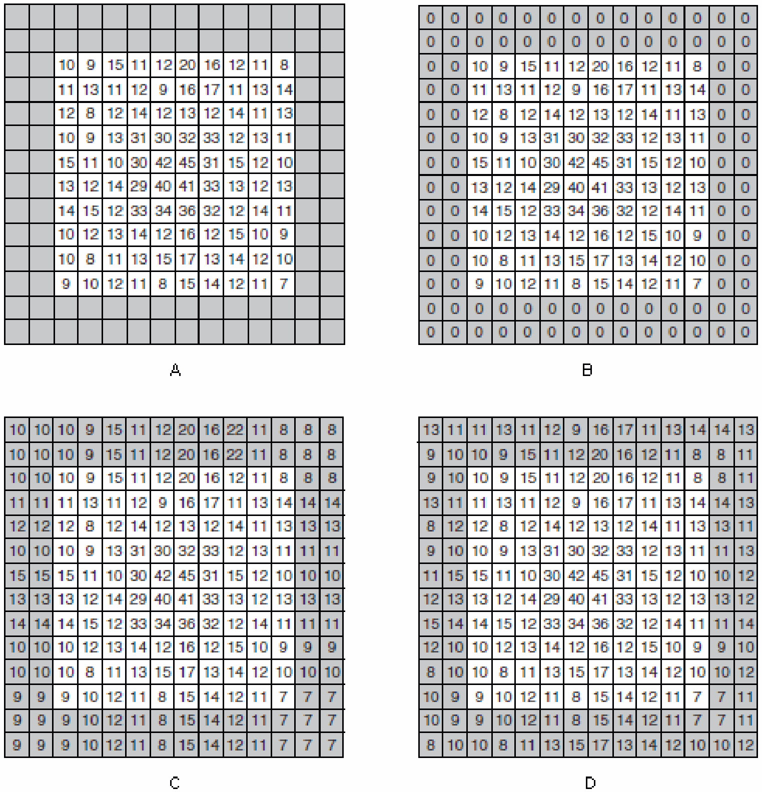


Image segmentation

$$\hat{y}_{ij} = f_\theta(x)_{ij}$$

- Segmentation = pixel-wise classification
- Definition: Label each pixel in the image with a category label
- Do not differentiate planes/instances, only care about pixels

The process of medical image segmentation can be divided into the following stages:

1. Obtain medical imaging data set
2. Preprocess the images
3. Use appropriate medical image segmentation method
4. Performance evaluation

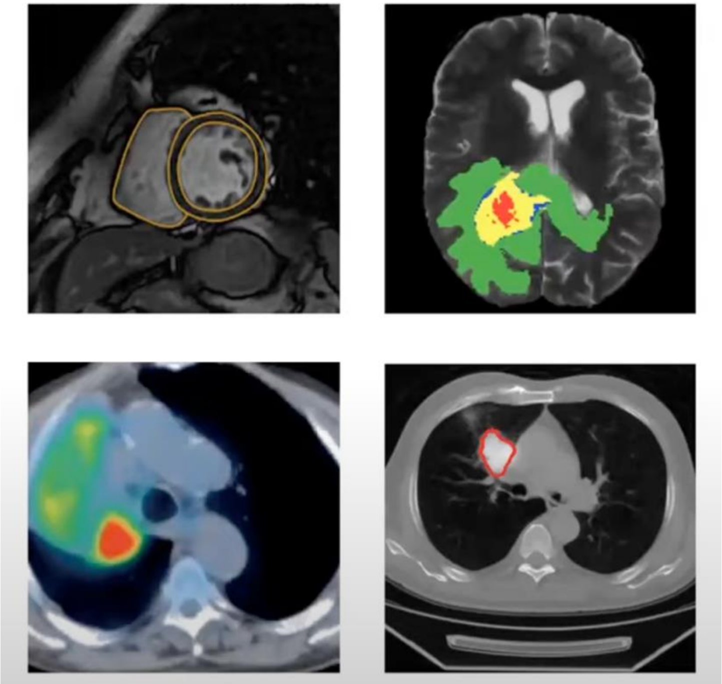
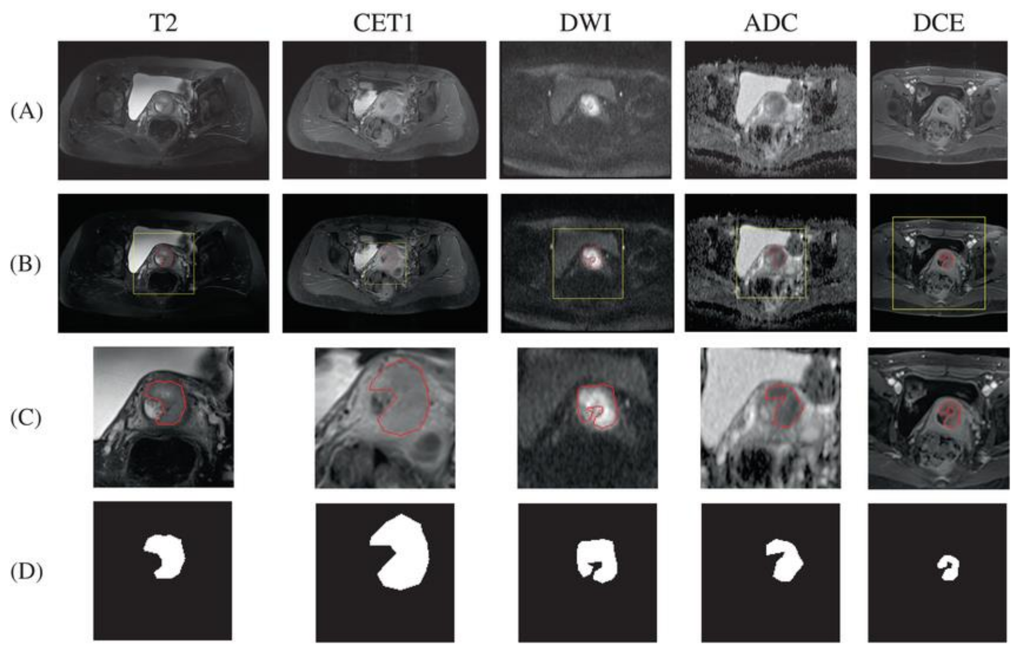


Image masks as ground truth for segmentation task

- Image Mask: $y \in \{0,1\}^{H \times W}$
- Used for:
  - Segmentation / Loss calculation / Performance evaluation



Examples of MRI and mask preprocessing.

(A) Raw MRI images;
(B) cervical cancer annotations drawn by clinicians.
(C) cropped MRI images overlapped with tumor annotations
(D) cropped binary masks with tumor regions

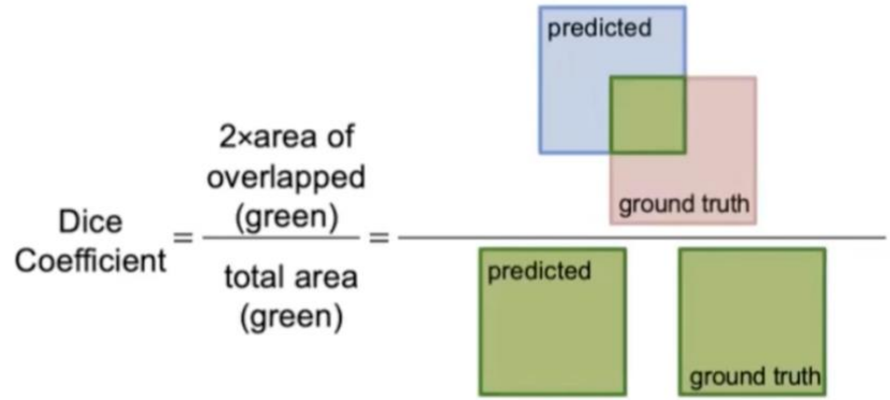
Image segmentation evaluation metrics

Evaluation metrics

- Dice is particularly useful when foreground objects are small relative to background.

Diced index = $\dfrac{2|X \cap Y|}{|X| + |Y|}$

$= \dfrac{2TP}{2TP + FP + FN}$



Jaccard Index $Jaccard(\mathrm{A},\mathrm{B}) = \dfrac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|}$ = TP / (TP + FP + FN)

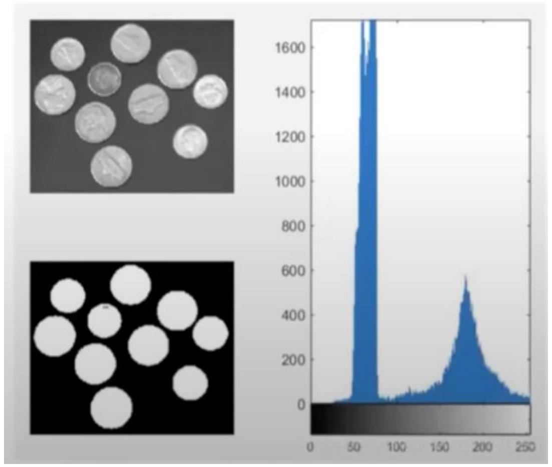
Image segmentation

Traditional methods

- Thresholding
- Edge detection
- Region growing

Limitations

- Sensitive to noise
- No global context
- Cannot learn high-level features



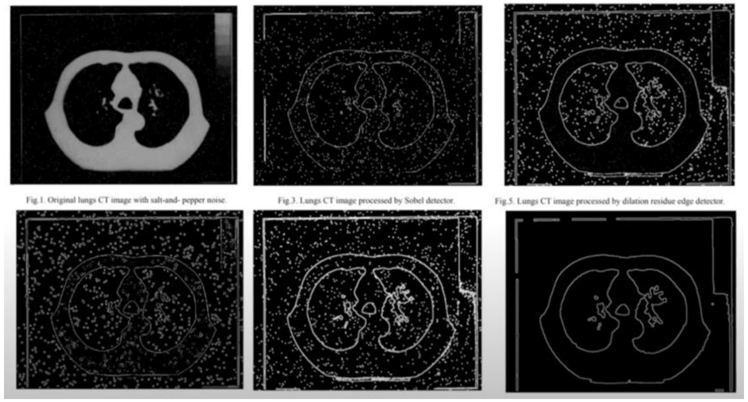
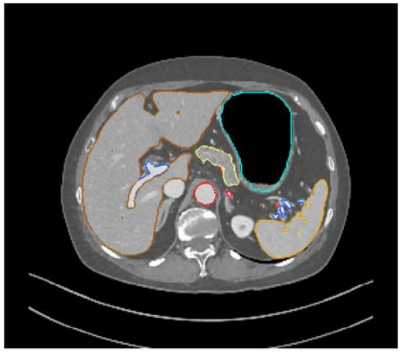
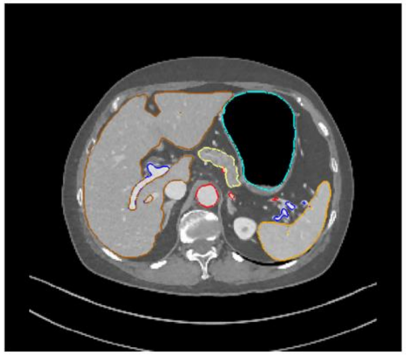


Image segmentation

- Deep learning based method - FCN



(a) Ground truth (axial)



(c) Segmentation (axial)

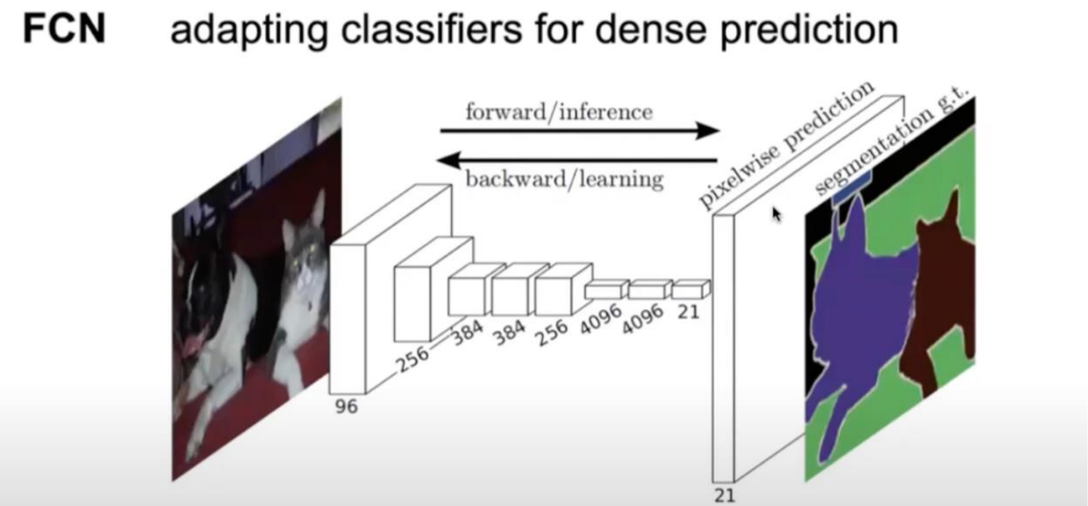
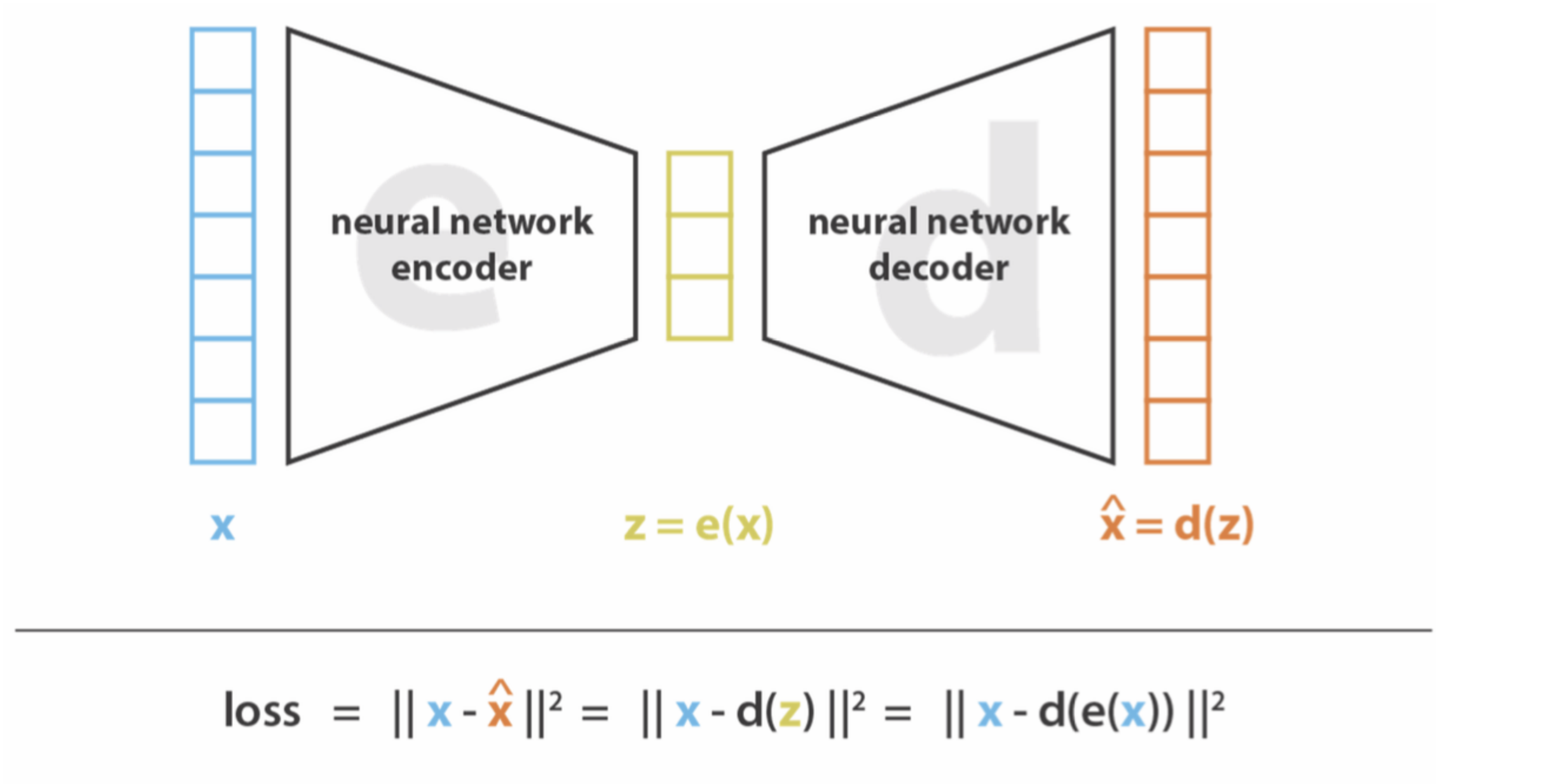


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

Roth, Holger R., et al. "Deep learning and its application to medical image segmentation." Medical Imaging Technology 36.2 (2018): 63-71.

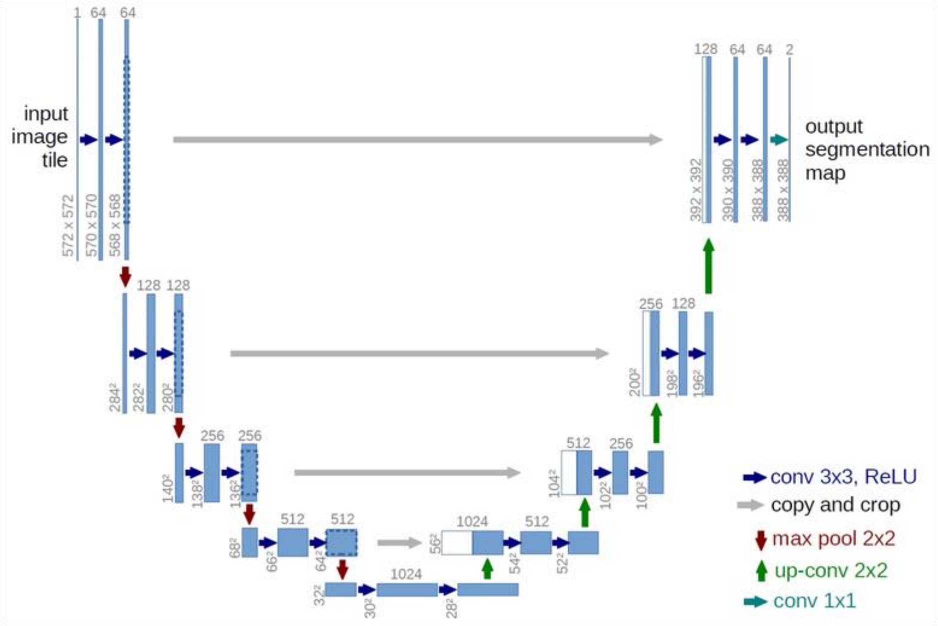
Encoder–Decoder Architecture

- The encoder learns a representation of the image.



Encoder–Decoder for Segmentation (U-Net)

U-Net's Encoder-Decoder Design Remains the Architectural Backbone of Medical Image Segmentation



Loss function:

$$Dice = \frac{2TP}{2TP + FP + FN}$$

Encoder Contracting path with convolution + max-pooling to extract hierarchical features.

Bottleneck Captures the most abstract, semantically rich representation of the image.

Decoder Expanding path with up-convolutions to restore spatial resolution.

Skip Connections Concatenate encoder features to decoder layers to preserve fine-grained spatial details.
Skip connections preserve high-resolution spatial information lost during pooling.

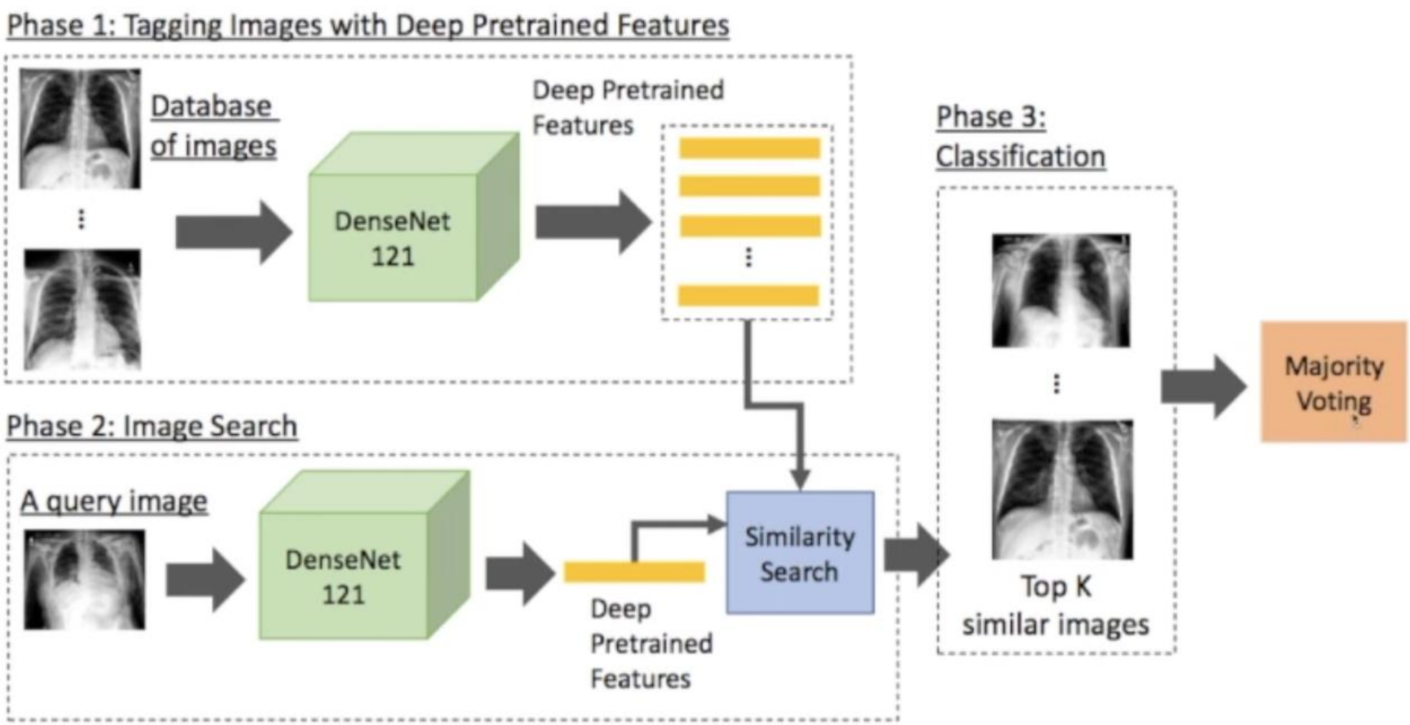
Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Encoder for Feature Representation

Instead of predicting masks, we extract features: $z = f_\theta(x)$

These embeddings can:

- Classify / Cluster / Retrieve similar images



Sze-To, A., Riasatian, A. & Tizhoosh, H.R. Searching for pneumothorax in x-ray images using autoencoded deep features. *Sci Rep* **11**, 9817 (2021). https://doi.org/10.1038/s41598-021-89194-4

Deterministic Prediction vs Probabilistic Modeling

Discriminative Models

- **Learn mapping: x → y**
- **Focus on decision boundaries**
- **Used for segmentation & classification**
- **Example: U-Net**

$\boxed{p(y \mid x)}$

Generative Models

- **Learn distribution of data**
- **Model what "normal" looks like**
- **Can generate or detect anomalies**
- **Example: VAE**

$\boxed{p(x) \text{ or } p(x \mid y)}$

Deterministic representation vs probabilistic representation.

Variational Autoencoder for image generation and abnormality detection

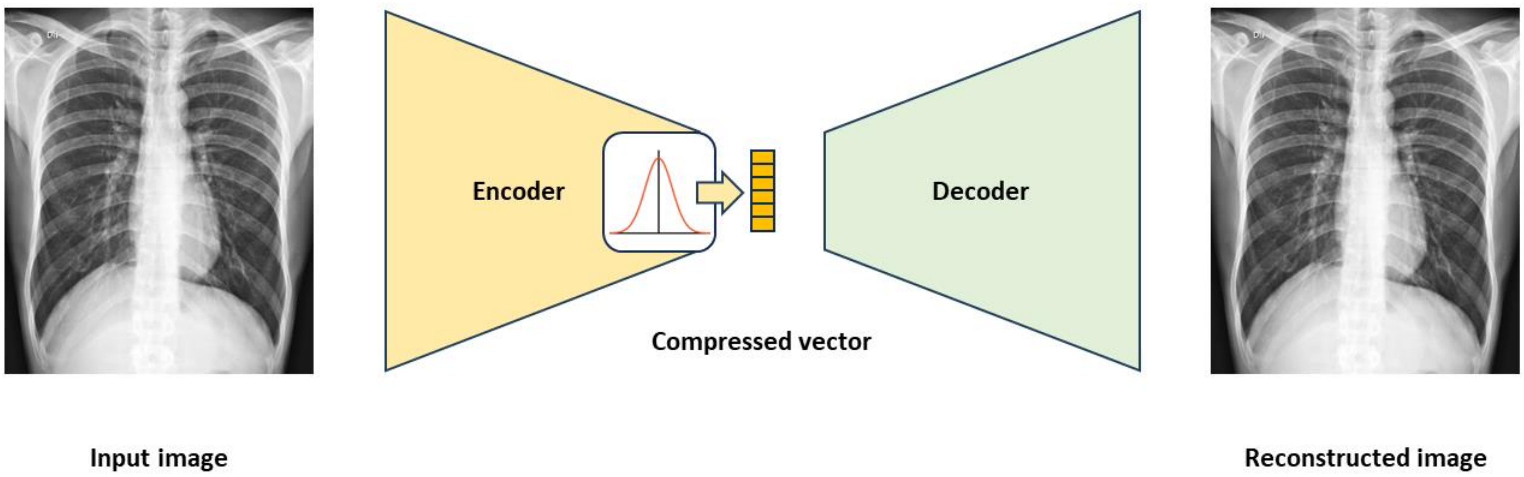
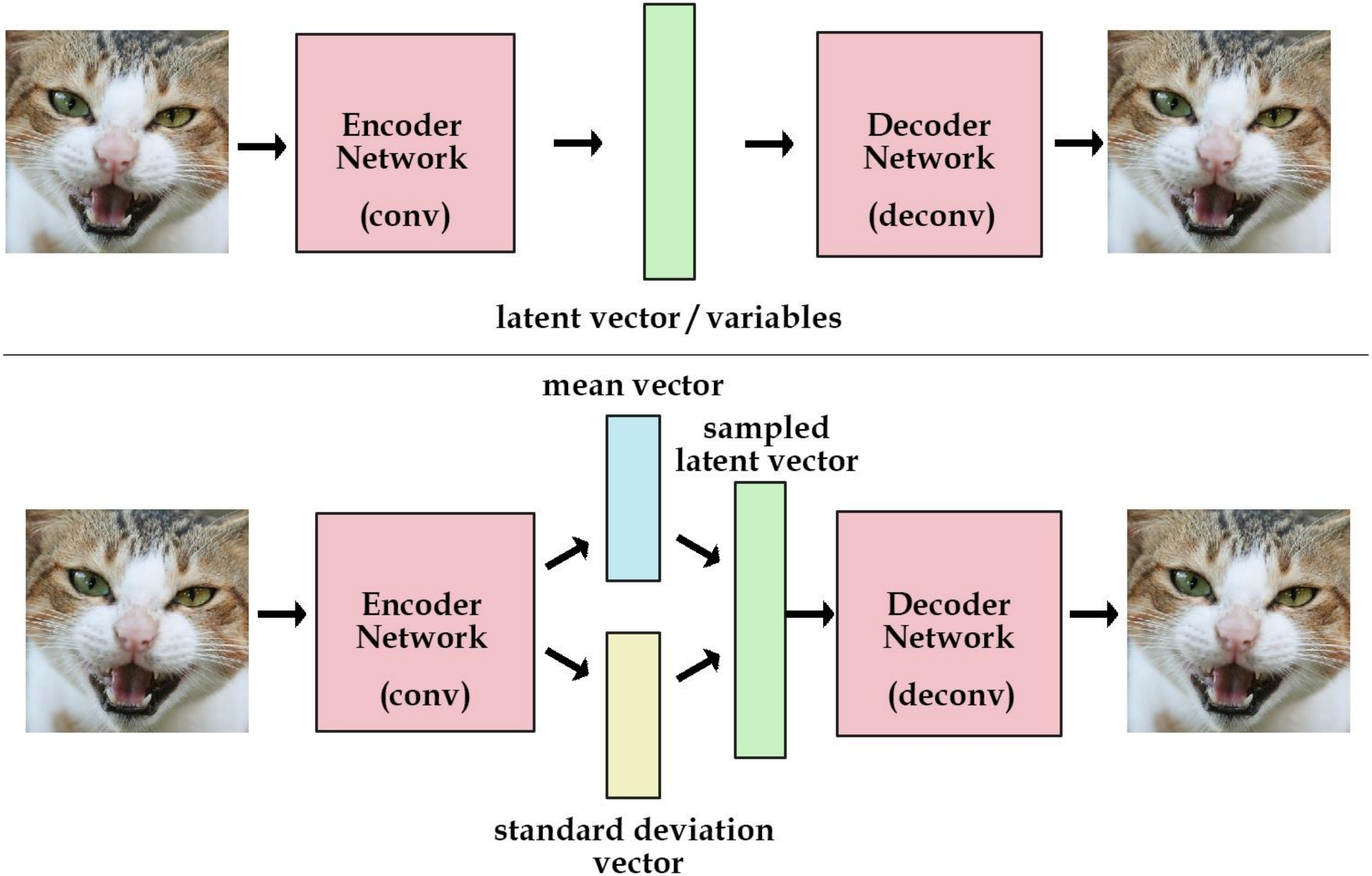


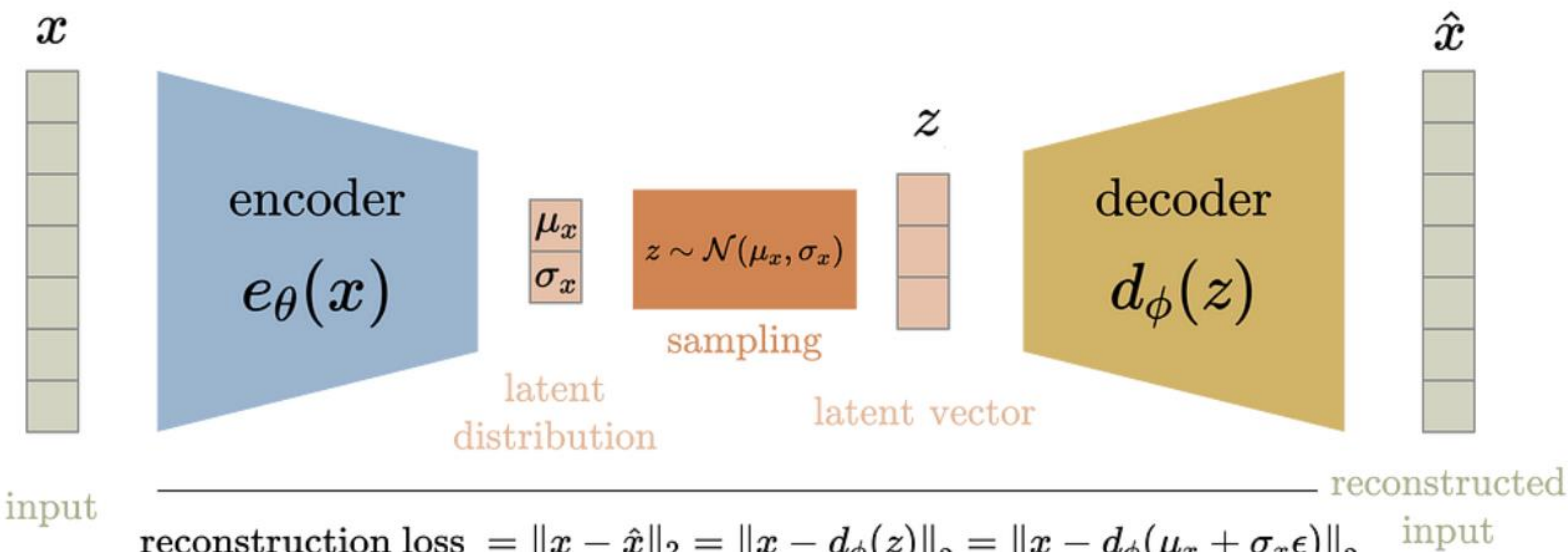
Image adopted from https://medium.com/@anthony.demeusy/ai-for-medical-imaging-detecting-anomalies-with-variational-autoencoders-d6fb37fdbc36

Autoencoder vs. Variational Autoencoder



https://kvfrans.com/variational-autoencoders-explained/

Variational Autoencoder explained



$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

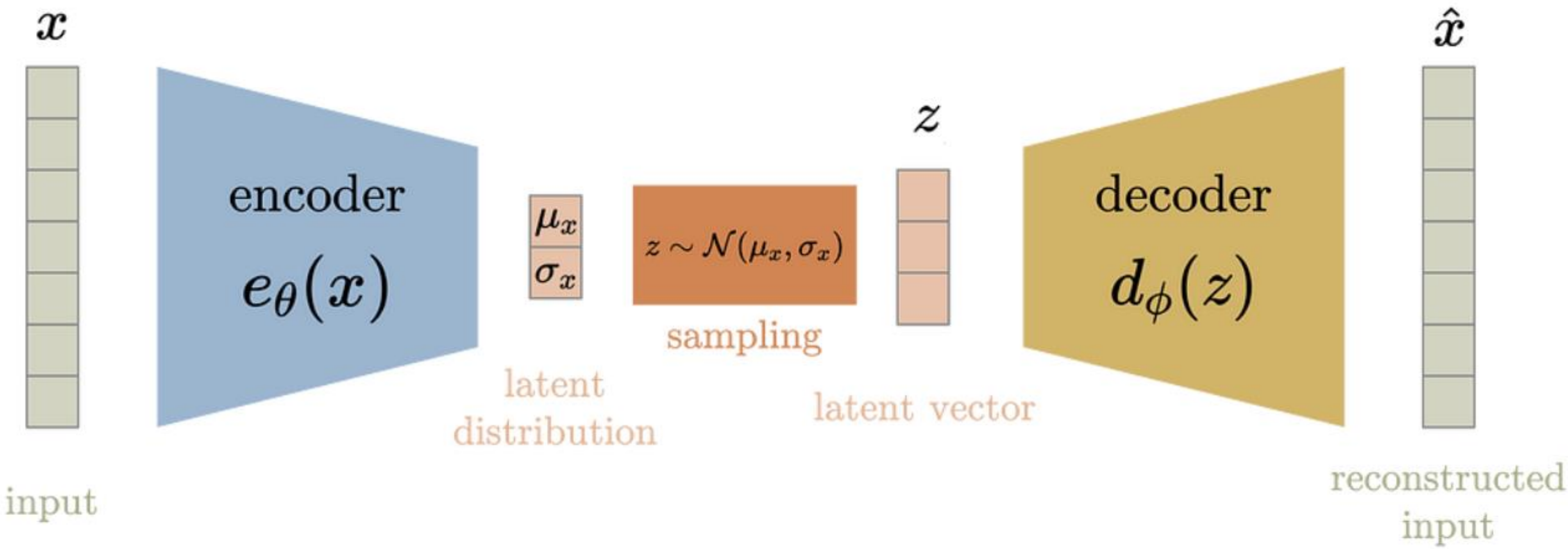
$$\textit{similarity loss} = \textit{KL Divergence} = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$\textit{loss} = \textit{reconstruction loss} + \textit{similarity loss}$$

- KL regularizes the latent space to follow a known prior, enabling sampling and interpolation.

Image adopted from https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2/

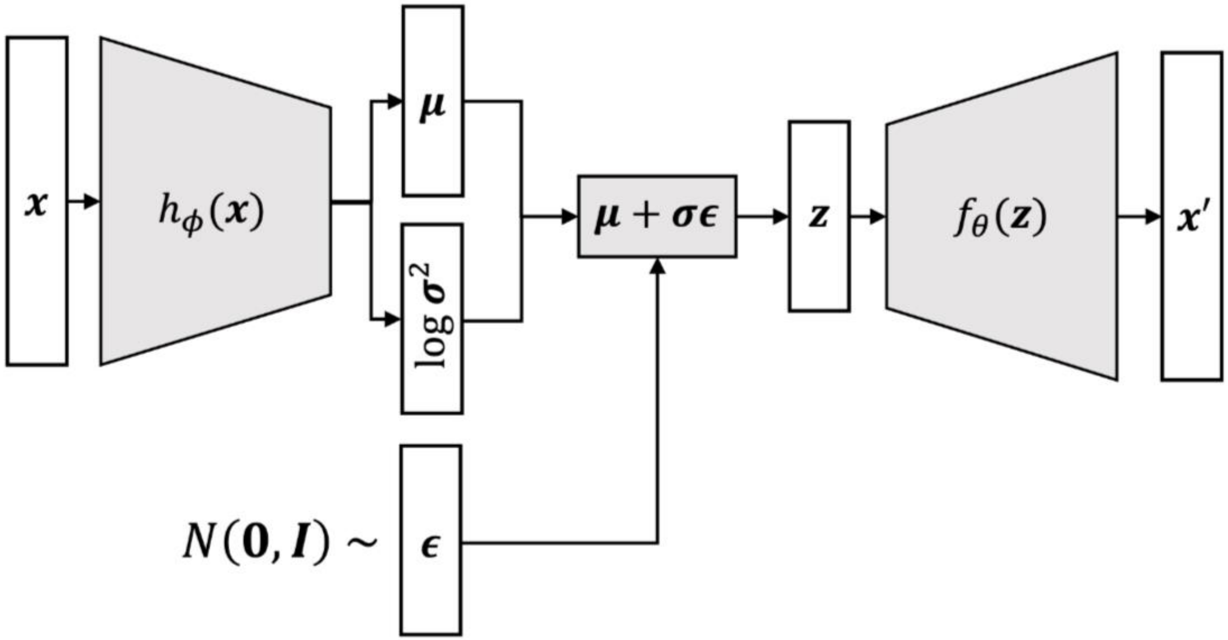
VAE – Reparameterization trick



The sampling step $z \sim N(\mu_x, \sigma_x)$ is a stochastic operation with no defined gradient, blocking backpropagation through the network.

Image adopted from https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2/

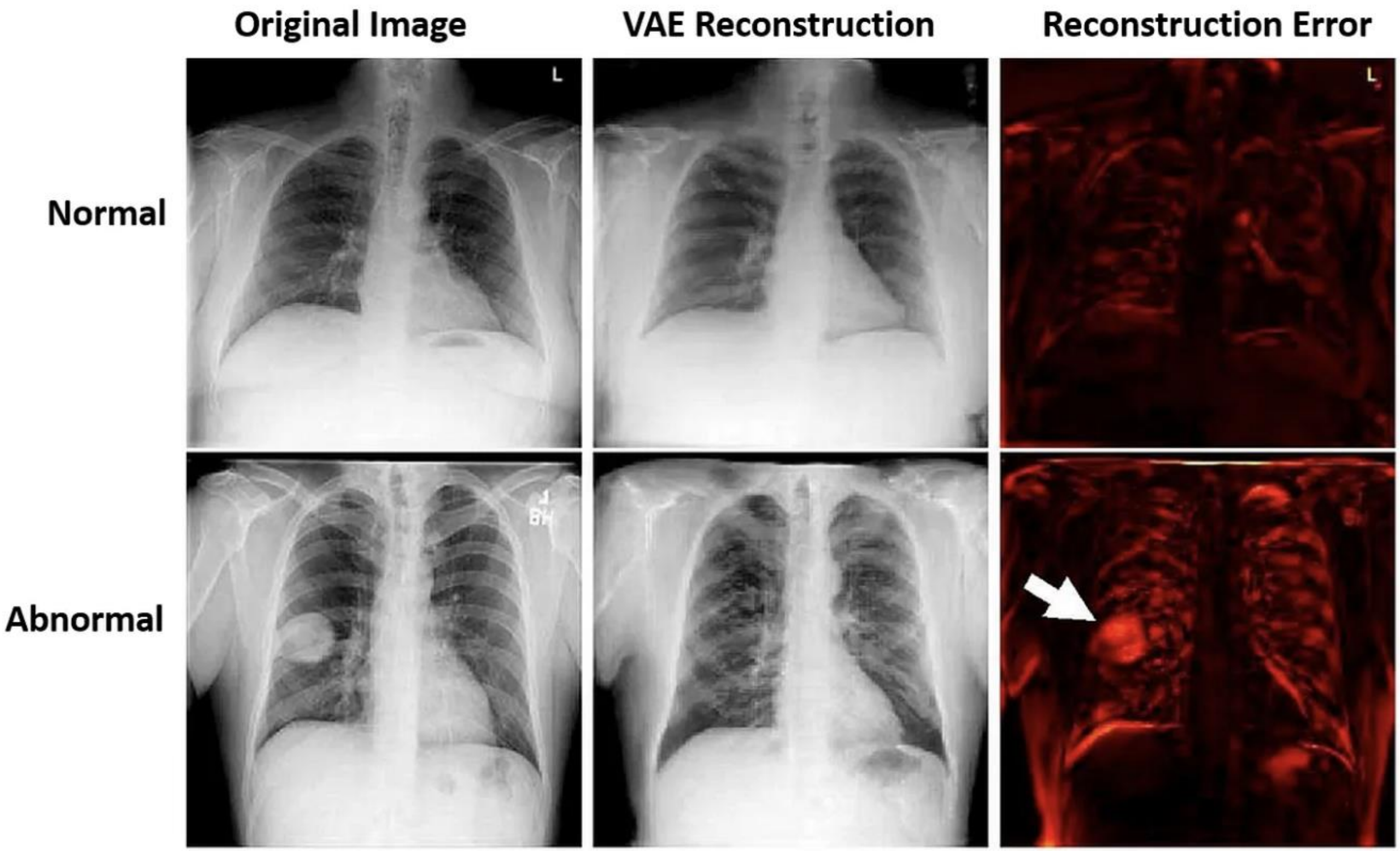
VAE – Reparameterization trick



The reparameterization trick moves randomness out of the network

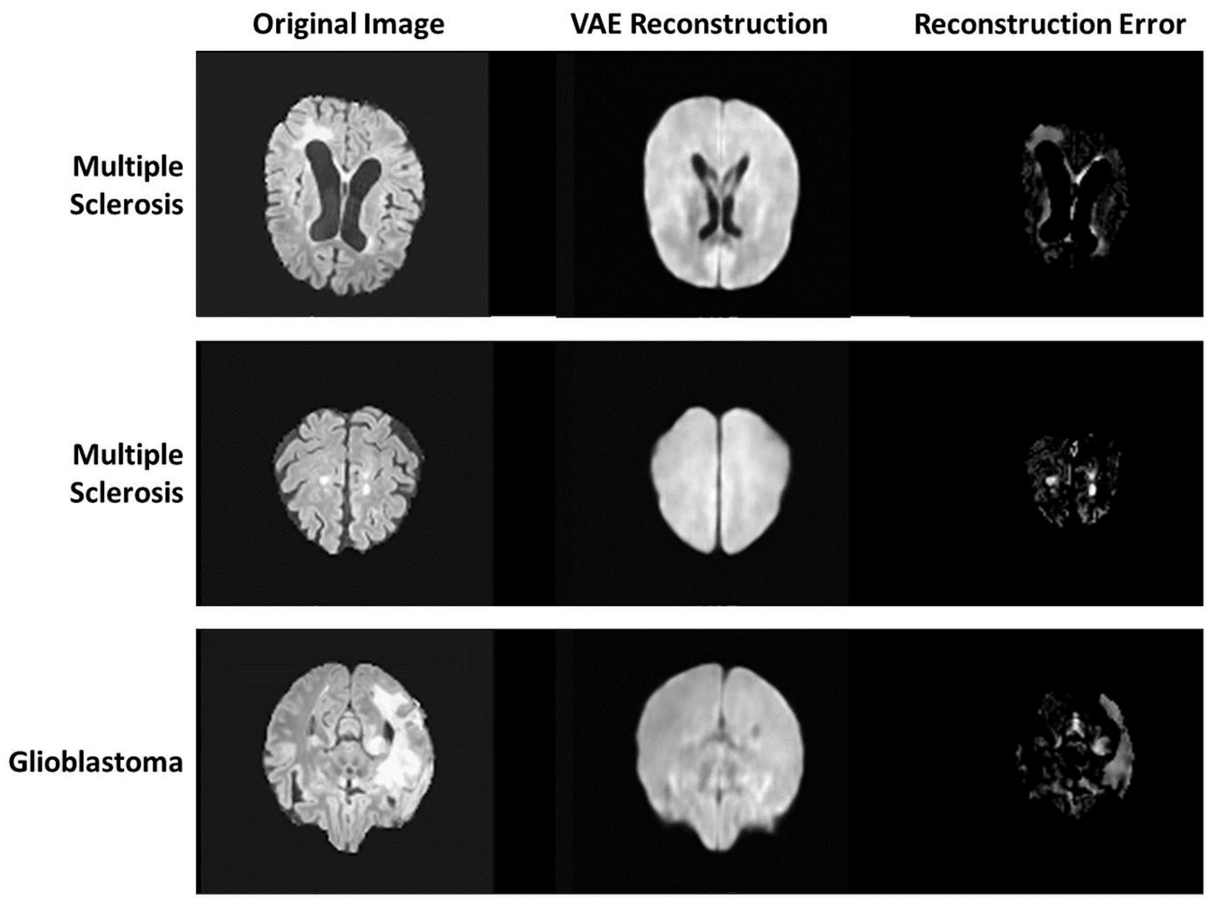
Instead of sampling z directly, we sample ε ~ N(0,I) externally and compute z = μ + σε, making the path through μ and σ fully differentiable.

Anomaly detection on chest X-ray using VAE



Nakao, Takahiro, et al. "Unsupervised Deep Anomaly Detection in Chest Radiographs." Journal of Digital Imaging, 8 Feb. 2021, https://doi.org/10.1007/s10278-020-00413-2. Accessed 10 Nov. 2024.

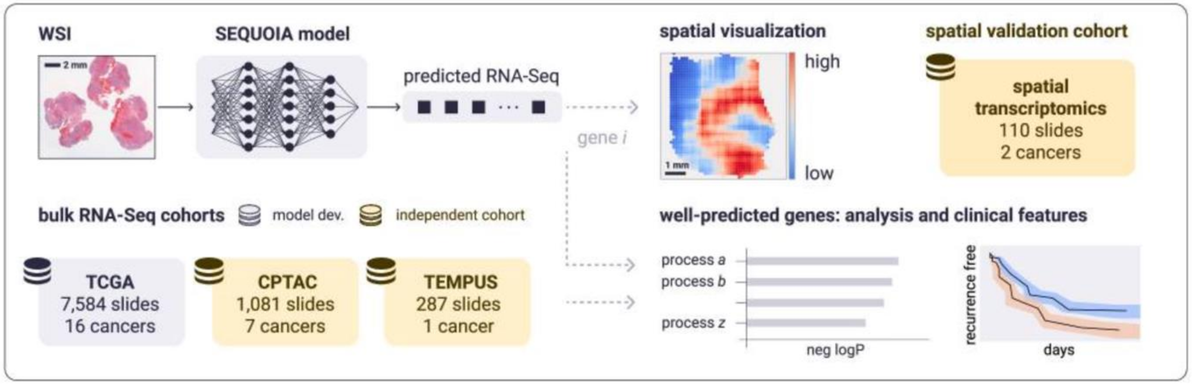
MRI, VAE reconstruction and reconstruction error



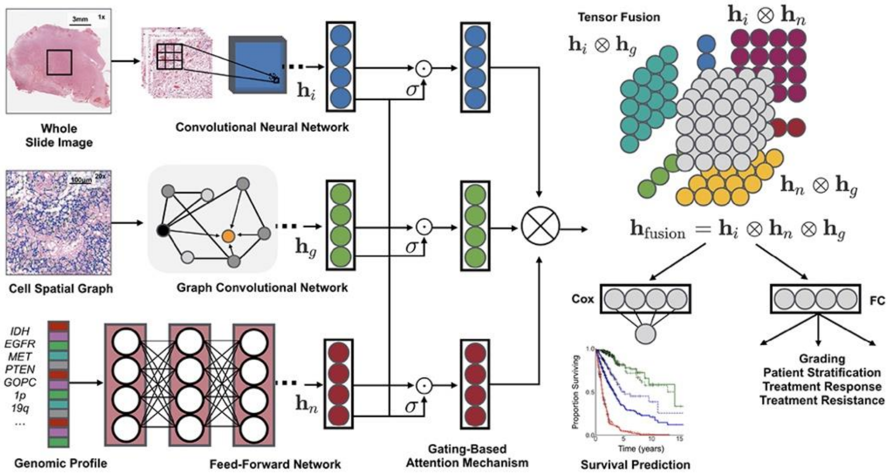
Baur, Christoph, et al. "Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study." *ArXiv:2004.03271 [Cs, Eess]*, 8 Apr. 2020, arxiv.org/abs/2004.03271. Accessed 10 Nov. 2024.

Multimodal AI: Integrating Imaging with Genomics

- Vision transformers can map whole-slide images to predicted RNA-seq profiles. [1]
- Deep models can predict gene expression, mutations, and pathway activity directly from whole-slide images.



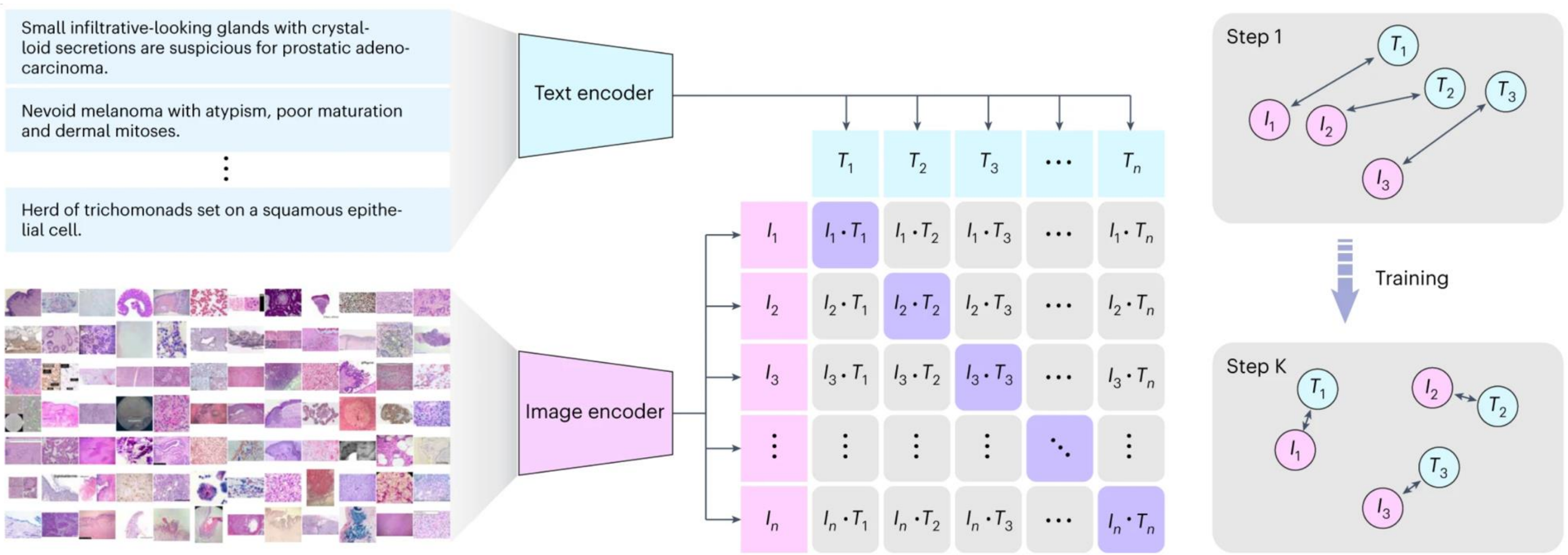
- Cross-modal fusion models integrate CNN image features with genomic embeddings. [2]
- Morphology encodes molecular state.



[1] Pizurica M. Digital profiling of gene expression from histology images with linearized attention. Nat Commun. 2024
[2] Richard C. Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. IEEE TransMed Imaging. 2023

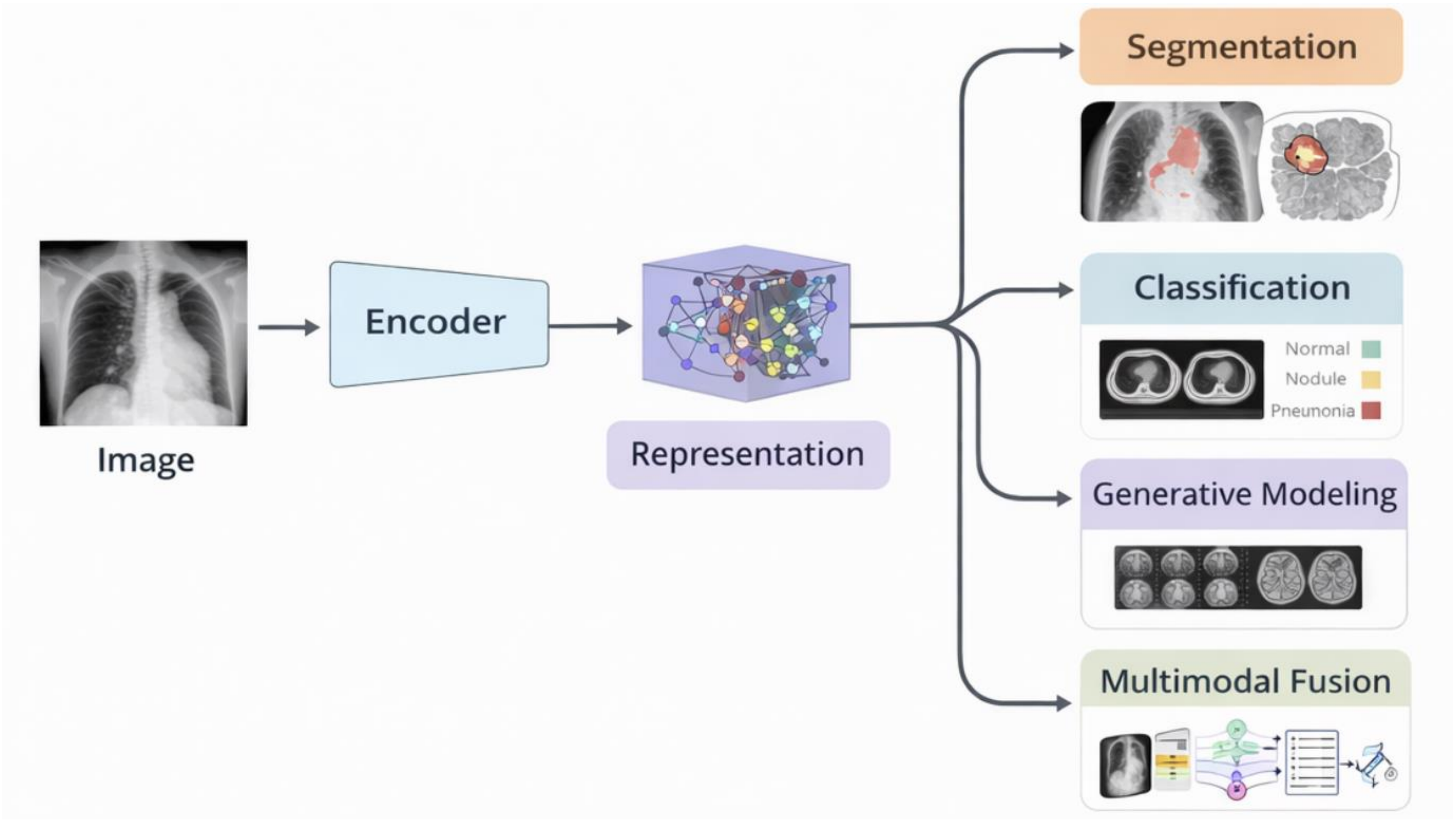
Multimodal AI: Integrating Imaging with Text

- Pathology text-image data pair from Twitter (X)
- Contrastive learning aligns image and text embeddings in a shared latent space.



Huang Z, Bianchi F, Yuksekgonul M, Montine TJ, Zou J. A visual-language foundation model for pathology image analysis using medical Twitter. Nat Med. 2023 Sep;29(9):2307-2316. doi: 10.1038/s41591-023-02504-3. Epub 2023 Aug 17. PMID: 37592105.

From Pixels to Probability to Multimodal Learning



Key Takeaways

- Medical images are high-dimensional tensors
- Segmentation = structured pixel prediction
- Latent space = learned representation of data structure.
- Discriminative models answer: *What is this?*
- Generative models answer: *What does this look like?*
- Multimodal AI connects morphology to molecular phenotype through shared representation learning.