Most Used Deep Learning Components (2025)

Activation Functions

Function	Typical Use Case	Key Property
ReLU	Hidden layers (CNNs, MLPs)	Fast, avoids vanishing grad
Leaky ReLU	When ReLU causes dead neurons	Small neg. slope for <0
GELU	Transformers, modern architectures	Smooth, probabilistic
Sigmoid	Binary classification output	Outputs [0,1]
Tanh	Recurrent networks, some hidden	Outputs [-1,1]

Loss Functions

Loss Function	Typical Use Case	Key Property
Cross-Entropy	Classification	Measures prob. divergence
MSE	Regression	Penalizes large errors
Binary CE	Binary classification	Log loss for two classes
Huber Loss	Robust regression	Less sensitive to outliers
KL Divergence	Probability distributions	Measures distribution diff

Optimizers

Optimizer	Typical Use Case	Key Property
Adam	Most deep learning tasks	Adaptive learning rate
AdamW	Modern architectures	Adam + weight decay fix
SGD (w/ momentum)	Traditional, some vision tasks	Simple, needs tuning
RMSprop	Recurrent networks	Adaptive per-parameter
Nadam	Combines Adam + Nesterov momentum	Smoother convergence

Metrics

Metric	Typical Use Case	Key Property
Accuracy	Classification (balanced data)	Intuitive, but misleading

Metric	Typical Use Case	Key Property
Precision/Recall	Imbalanced classification	Focus on positives/negatives
F1 Score	Imbalanced classification	Harmonic mean of P/R
AUC-ROC	Probabilistic classification	Measures ranking ability
IoU (mIoU)	Segmentation	Overlap quality
MSE/RMSE	Regression	Error magnitude

Trends (2025)

- **Activation:** GELU and Swish variants are gaining in transformers.
- **Optimizers:** AdamW and Lion (new adaptive optimizer) are popular for large models.
- **Loss:** Label smoothing and focal loss for robustness.
- **Metrics:** AUC-PR and custom business metrics for real-world impact.