ChimeraLM Architecture

Legend:
- **B** = batch size
- **L** = sequence length

Input
Tokenized DNA/RNA
[B, L]

[B, L] →

Hyena Operator
Convolution

[B, L, 256] →

Hidden States
[B, L, 256]

[B, L, 256] →

Attention Pooling
- Linear(256→128)
- GELU
- Linear(128→1)
- Softmax
- Weighted Sum

[B, 256] →

Pooled Output
[B, 256]

[B, 256] →

Classification
- Linear(256→512)
- GELU + Dropout
- Linear(512→512)
- GELU + Dropout
- ResidualBlock

[B, 512] →

Output
Linear(512→2)
[B, 2]