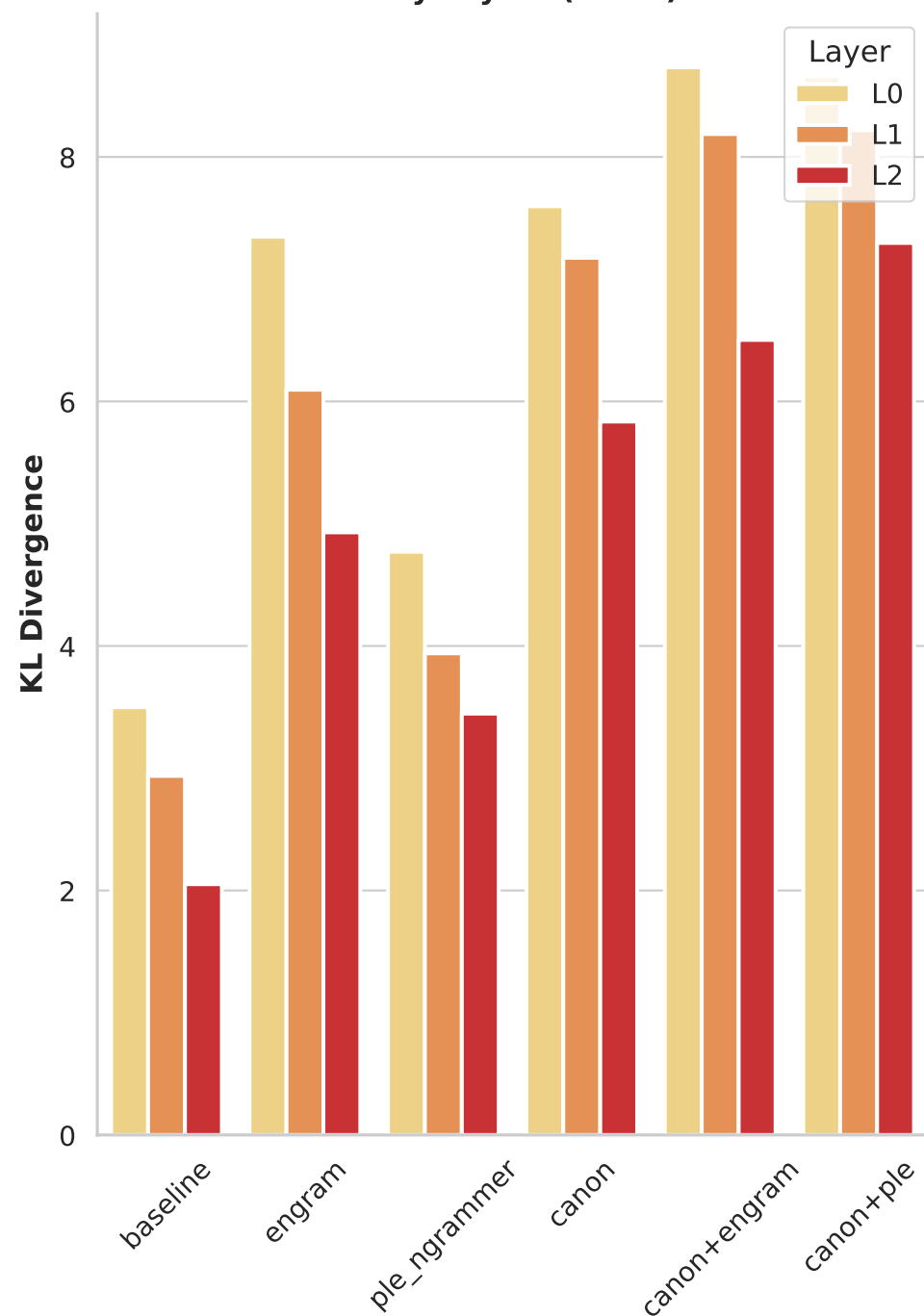Prediction Convergence: Early vs Late Layers