



Filter Characteristics

Filter size: 6×1
Filter stride: 1

27 Channels,
Where one
for each
char

Compute
Multi-output
multi-categorical
loss

3 Char
context
autoregressive
model

Rather than break sentences into multiple training examples when training an autoregressive MLP, can more efficiently train an equivalent model using convolutional layers