

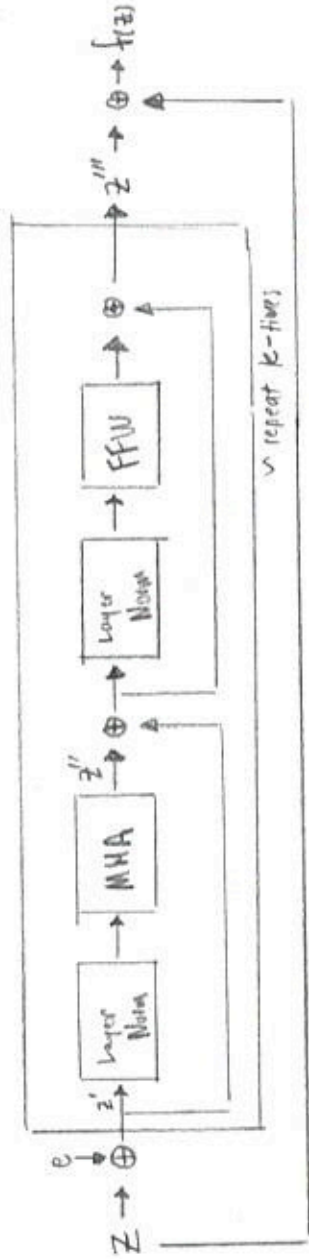
SepFormer modified with Persistence

SepFormer Baseline



repeat N -times

"SepFormer Block"



"Transformer block"

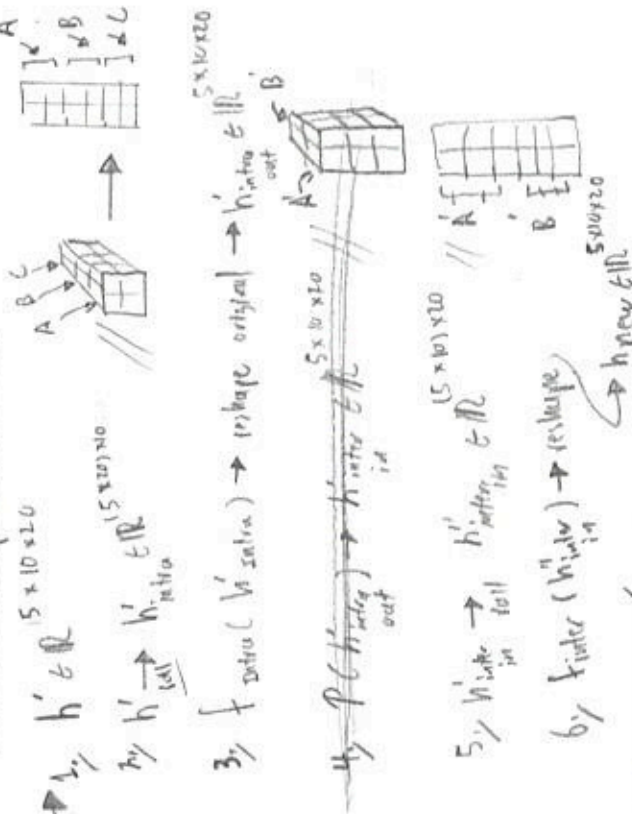
SepFormer Equations

- $h' \in \mathbb{R}^{F \times L \times N_C}$
- L , length of each chunk
- N_C , resulting number of chunks
- $h'' \in \mathbb{R}^{F \times L \times N_C}$

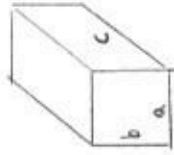
$$h'' = f_{inter} (P(f_{intra}(h')))$$

- Intra T \sim ind dim of h' , (short)
- Inter T \sim ind dim of $P(h')$ \sim $F \times N_C \times L$, (long)

Dim. Soudly Check

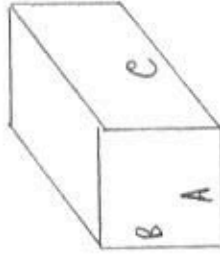


2. Chop Shop



$$\text{Latent} := \mathcal{L}_{a \times b \times c}$$

$$\text{Byte Array} := B_{A \times B \times C}$$



$$\text{Such: } \begin{aligned} a &\leq A \\ b &\leq B \\ c &\leq C \end{aligned}$$

$$\text{Let, } h_c = B_{(F \times C \times N_c) \times (A \times B \times C)}$$

[Long Cross Weights $\in W_{a \times c}^{a_L}, W_{c \times c}^{k_L}, W_{c \times c}^{V_L}$]
 [Short Cross Weights $\in W_{b \times b}^{a_S}, W_{b \times b}^{k_S}, W_{b \times b}^{V_S}$]

START

"Short Cross Attention"

$$\begin{aligned} \mathcal{L}_{a \times b \times c}^S &\xrightarrow{P} \tilde{\mathcal{L}}_{(a \times c) \times b}^S \xrightarrow{P} A_{(a \times c) \times b}^S \xrightarrow{P} A_{a \times b \times c}^S \\ B_{A \times B \times C} &\xrightarrow{P} B_{(A \times C) \times B}^S \xrightarrow{P} B_{(A \times B) \times C}^S \end{aligned}$$

$$N_{a \times b \times c}^S \xrightarrow{P} \tilde{\mathcal{L}}_{a \times b \times c}^S$$

Short Latent Transformation

"Long Cross Attention"

$$\tilde{\mathcal{L}}_{a \times b \times c}^L \xrightarrow{P} \tilde{\mathcal{L}}_{(a \times b) \times c}^L$$

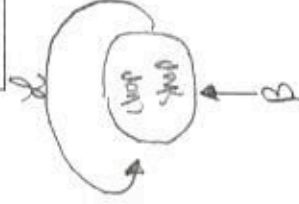
$$B_{A \times B \times C} \xrightarrow{P} B_{(A \times B) \times C}^L$$

$$\tilde{\mathcal{L}}_{(a \times b) \times c}^L \xrightarrow{P} A_{a \times b \times c}^L$$

END

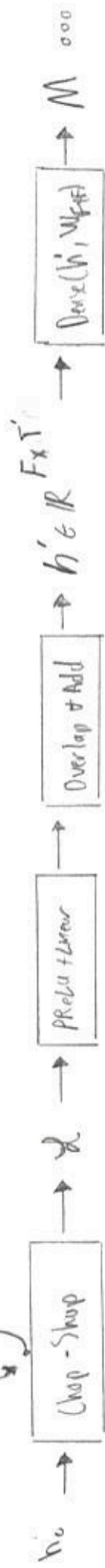
$$000 \xrightarrow{P} A_{a \times b \times c}^L \xrightarrow{P} A_{a \times b \times c}^{L'} \xrightarrow{P} \mathcal{L}_{a \times b \times c}^{new}$$

Long Latent Transformation



3. Decoder

$\times K$

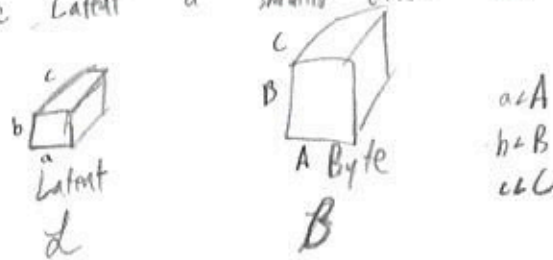


$$000 \rightarrow M \odot h_2 \rightarrow \text{Conv 1D Transpose} (M \odot h_2) \rightarrow \hat{S}^{T \times \# \text{ sources}}$$

Security Check: dim of chunk-sharp VERSION 2

Problem: d is smaller than x

Idea 1: Make Latent a smaller chunk size than byte chunk



Short Pass

$$\begin{aligned}
 &L_{a \times b \times c} \rightarrow P \rightarrow L_{(a \times c) \times b}^s \rightarrow L_{(a \times c) \times b}^s W_{b \times b}^{Q_s} \rightarrow Q_{(a \times c) \times b}^s \\
 &B_{A \times B \times C} \rightarrow P \rightarrow B_{(A \times C) \times B}^s \rightarrow \begin{cases} B_{(A \times C) \times B}^s W_{B \times b}^{K_s} \rightarrow K_{(A \times C) \times b}^s \\ B_{(A \times C) \times B}^s W_{B \times b}^{V_s} \rightarrow V_{(A \times C) \times b}^s \end{cases} \rightarrow \text{softmax} \left(\frac{Q_{(a \times c) \times b}^s K_{b \times (A \times C)}^{sT}}{\sqrt{d_b}} \right) V^s \\
 &\quad = A_{(a \times c) \times b}^s \\
 &\quad \downarrow P \\
 &\quad \downarrow P \\
 &\quad A_{a \times b \times c}^s
 \end{aligned}$$

Long Pass

✓ w.l.o.g with $\{W_{c \times c}^{Q_L}, W_{c \times c}^{K_L}, W_{c \times c}^{V_L}\}$

$$L_{a \times b \times c} \rightarrow P \rightarrow L_{a \times b \times c}^L$$

$$B_{A \times B \times C} \rightarrow P \rightarrow B_{(A \times B) \times C}^L$$