

# Attention based models in End-to-End ASR

Exploration of Attention in ESPNET toolkit

---

Shreekantha Nadig

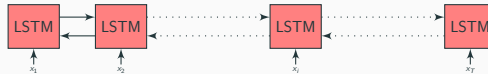
November 21, 2018

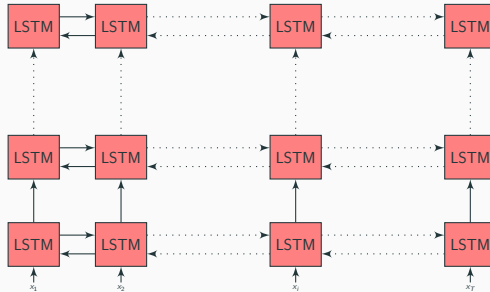
International Institute of Information Technology - Bangalore

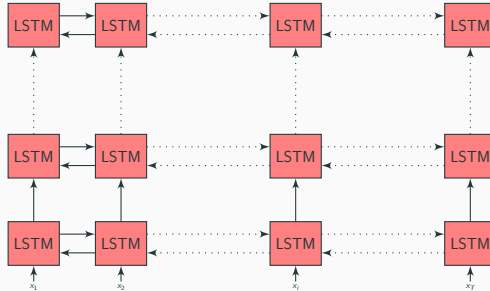
1. Introduction
2. No Attention [Equal Attention?]
3. Dot product Attention
4. Additive Attention
5. Location Aware Attention
6. 2D Location Aware Attention
7. Location Aware Recurrent Attention
8. Coverage mechanism Attention
9. Coverage mechanism location aware Attention
10. Multi-Head Attention

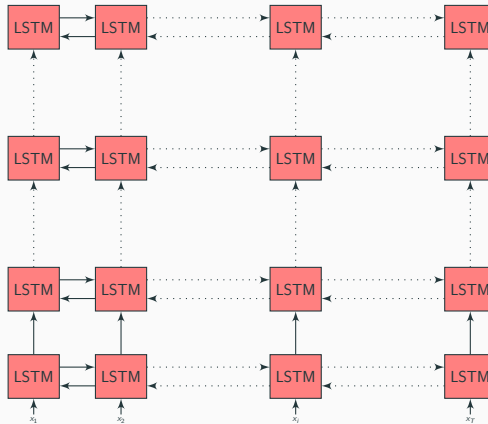
# Introduction

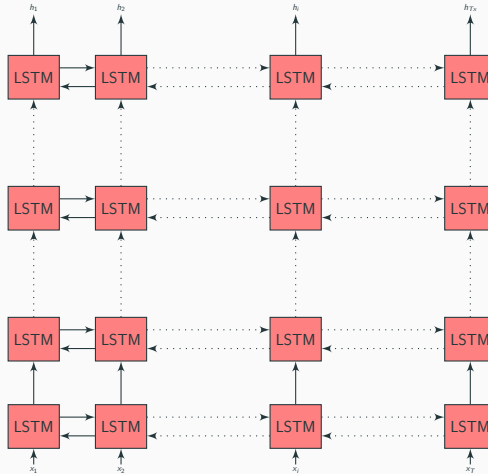
---





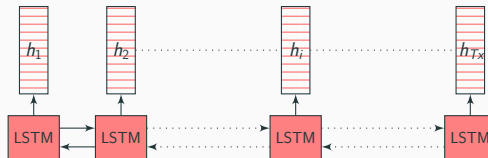




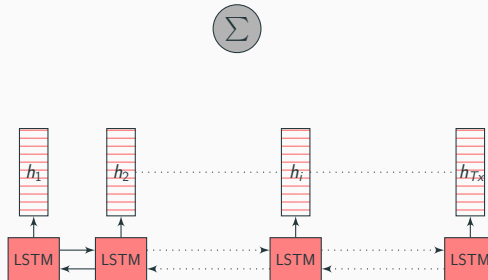


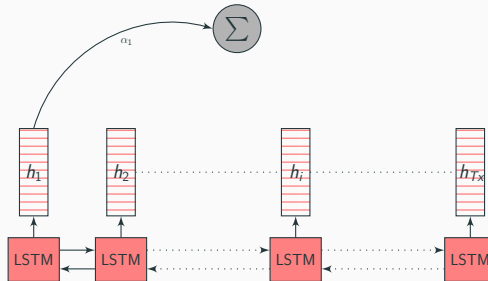


# Introduction to Attention

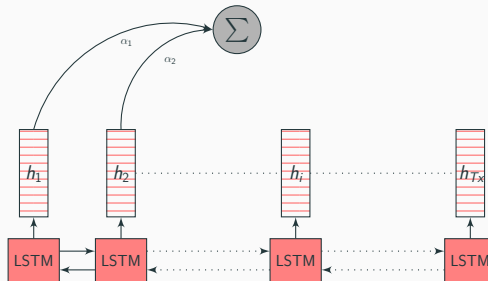


# Introduction to Attention

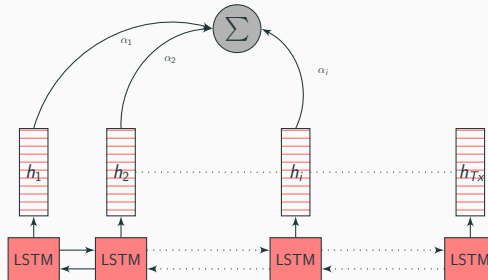


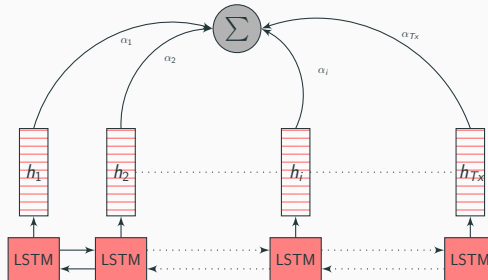


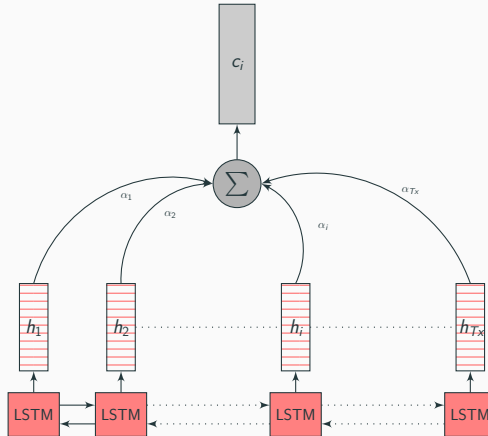
# Introduction to Attention



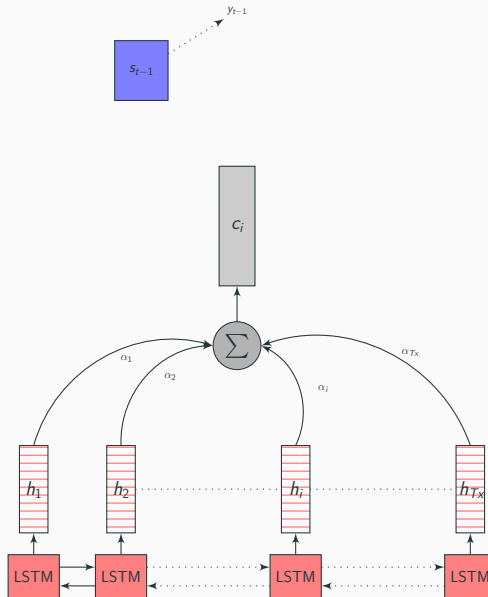
# Introduction to Attention





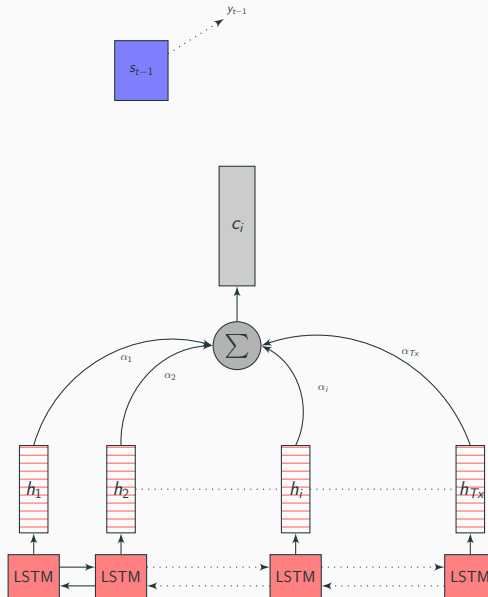


# Introduction to Attention

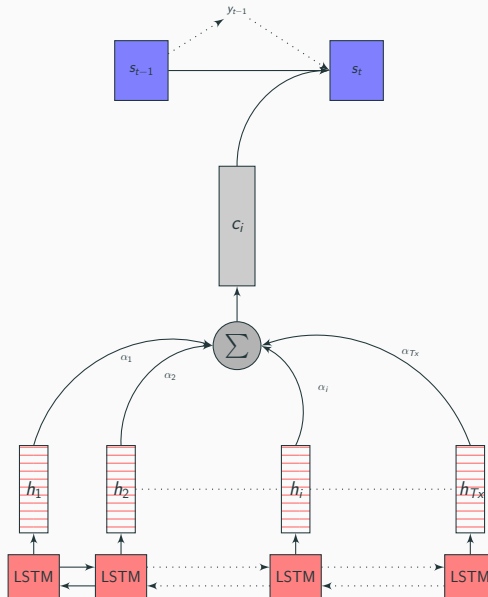


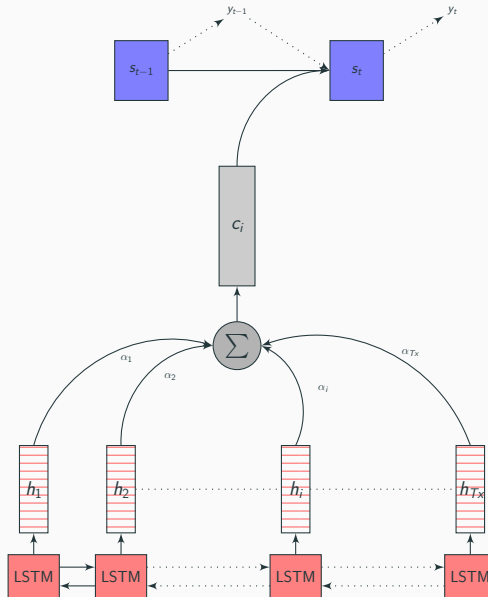


# Introduction to Attention



# Introduction to Attention

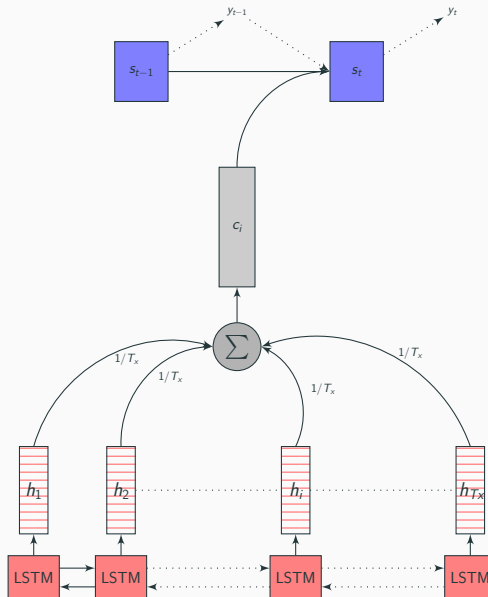


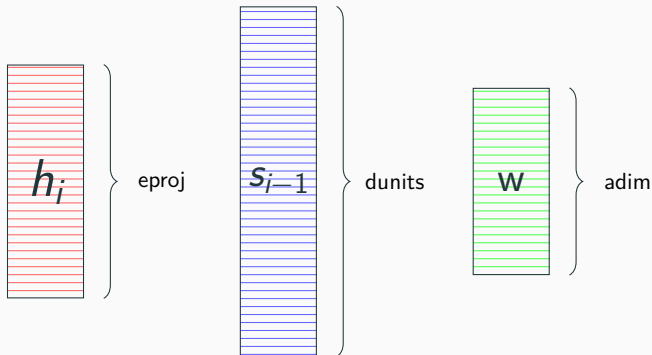


**No Attention [Equal Attention?]**

---

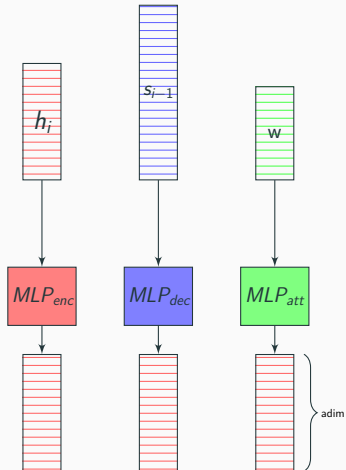
# No Attention [Equal Attention?]





Mostly  $eproj \neq dunits \neq adim$

# Matching the dimensions of representations



# Dot product Attention

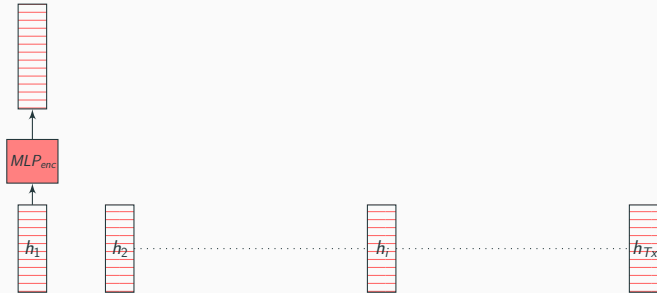
---



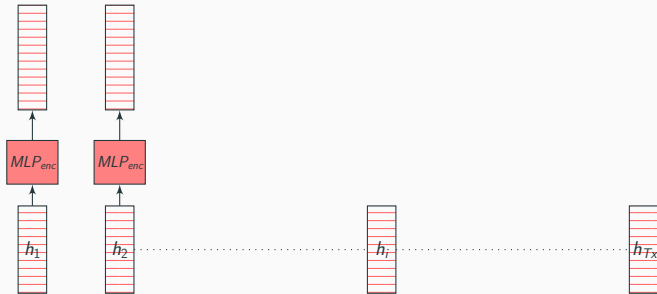
# Dot product Attention



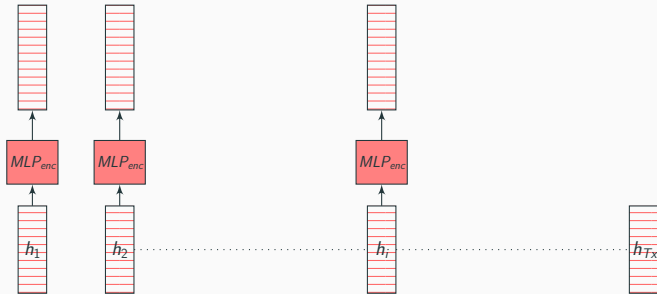
# Dot product Attention



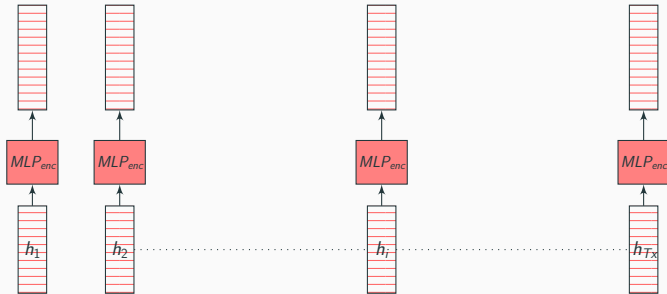
# Dot product Attention



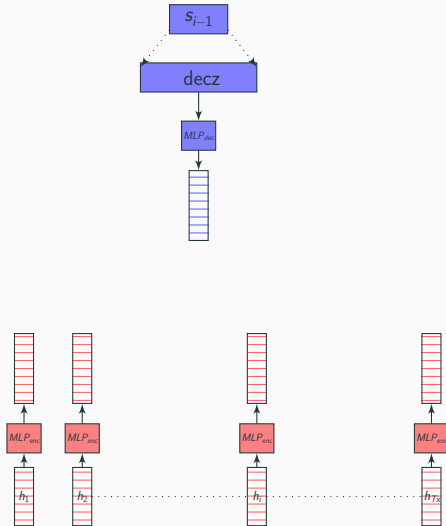
# Dot product Attention



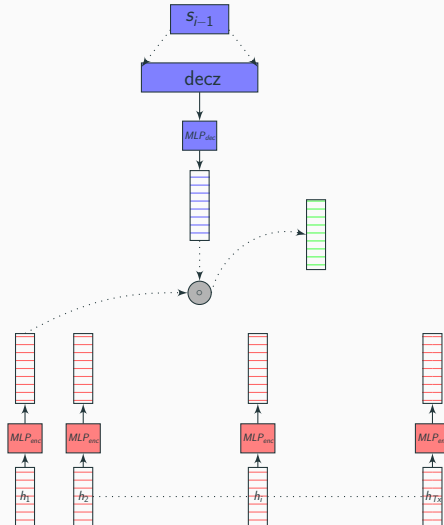
# Dot product Attention

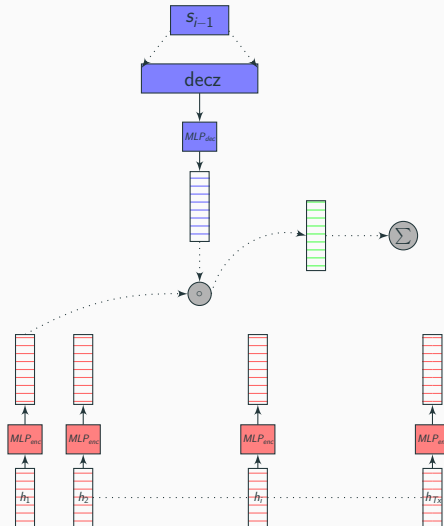


# Dot product Attention



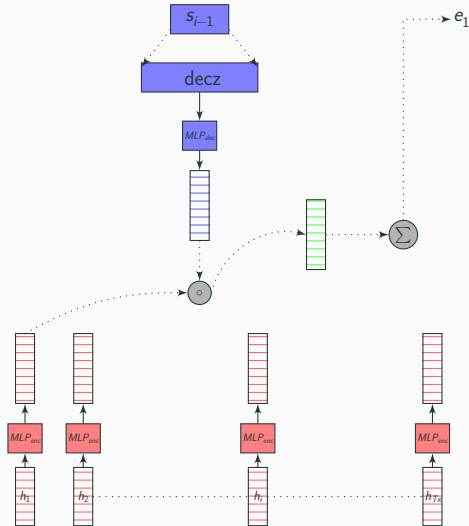
# Dot product Attention





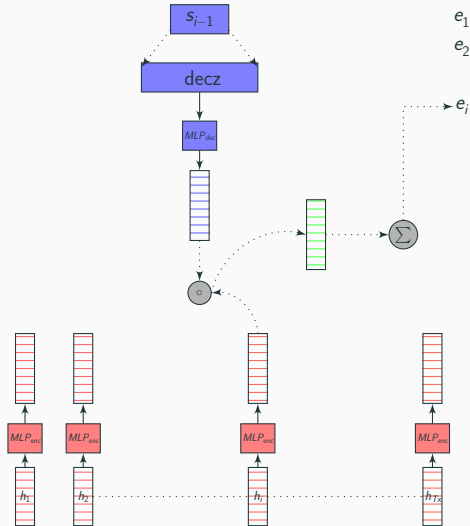


# Dot product Attention

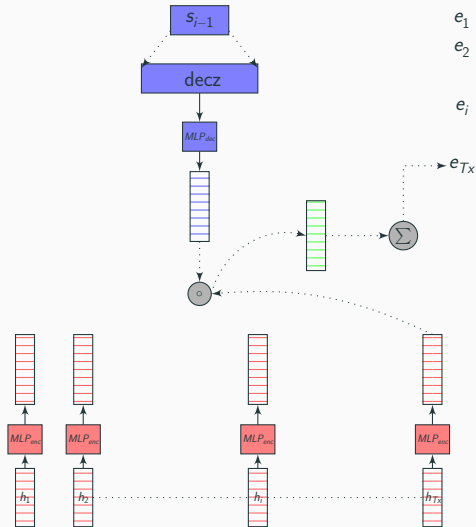




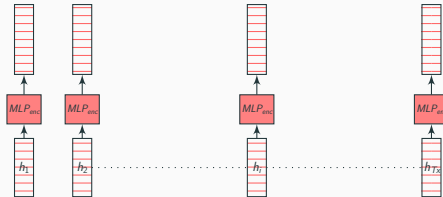
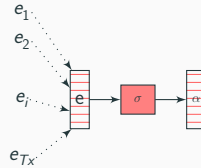
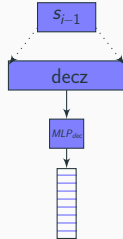
# Dot product Attention



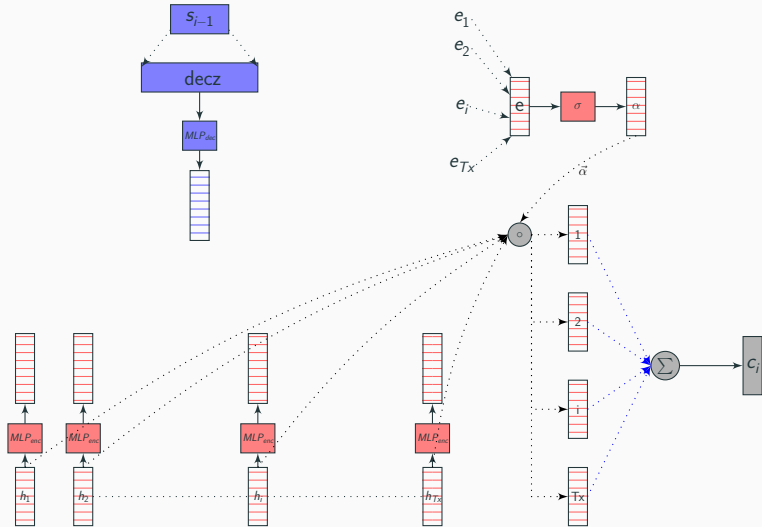
# Dot product Attention



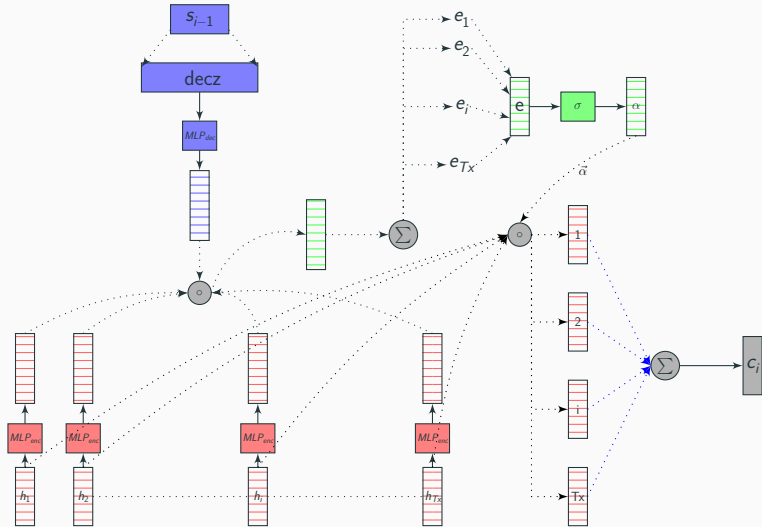
# Dot product Attention



# Dot product Attention



# Dot product Attention - Full picture

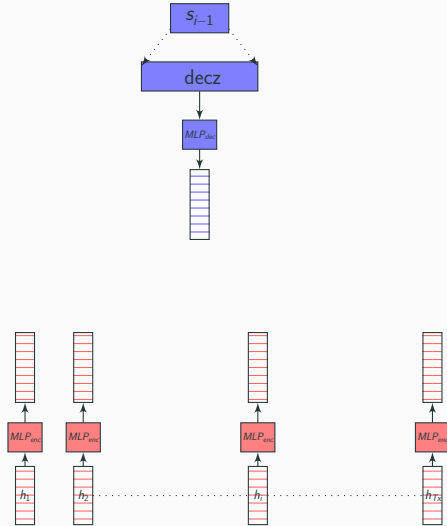


# Additive Attention

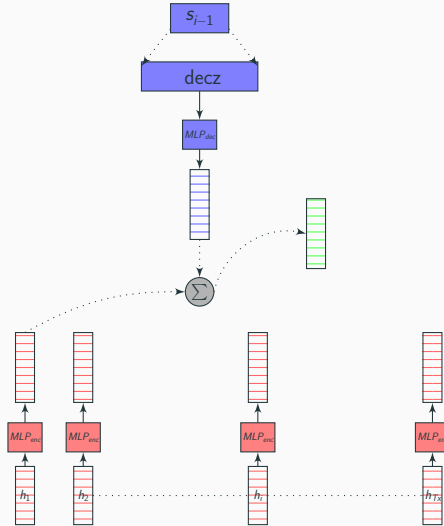
---

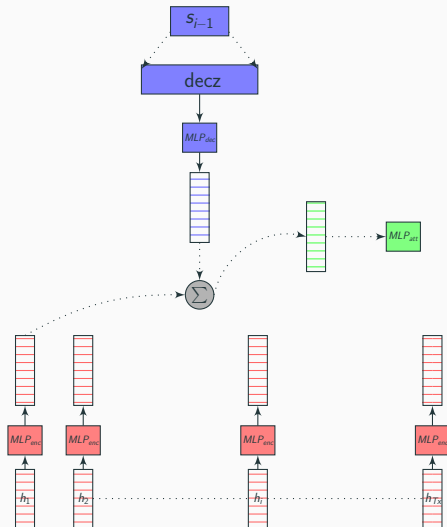


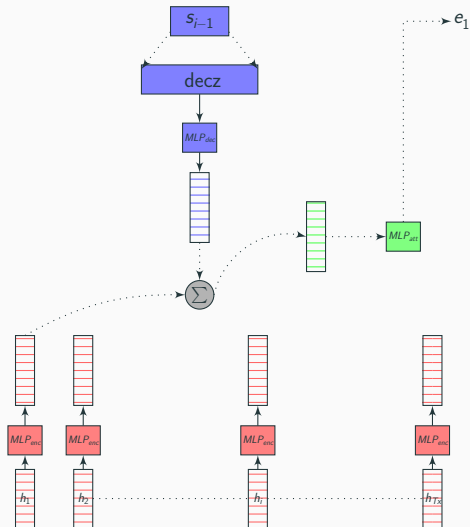
# Additive Attention

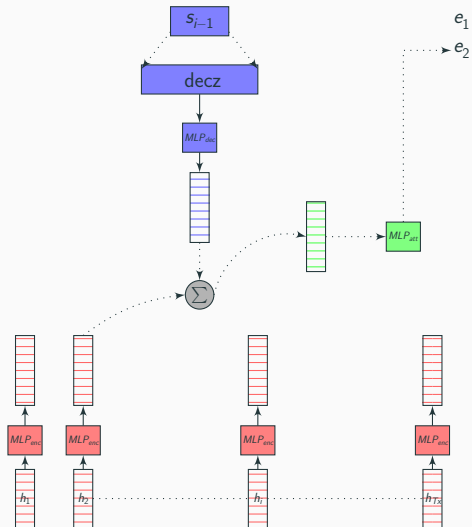


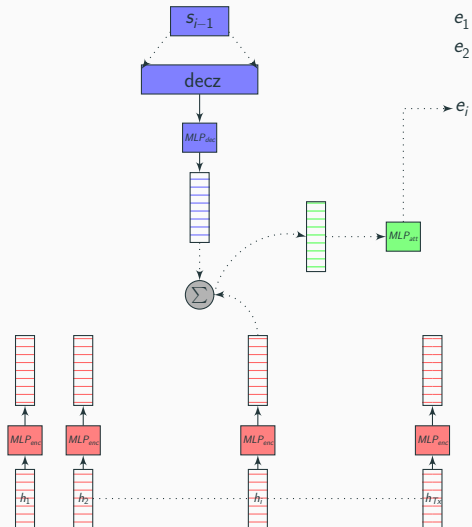
# Additive Attention

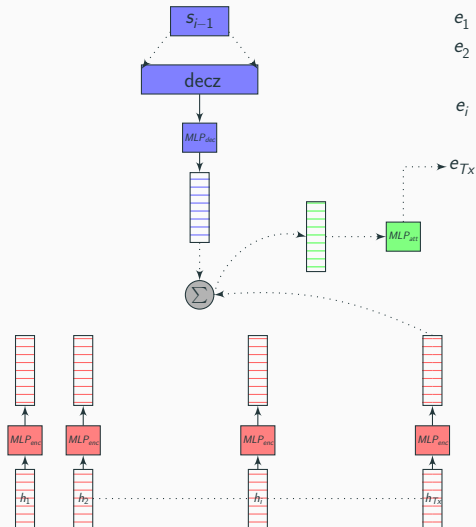


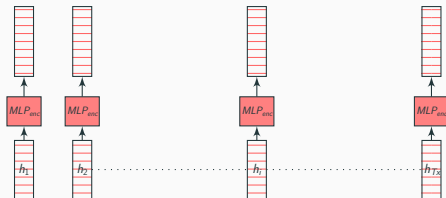
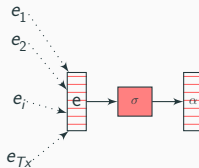
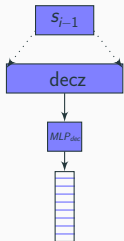






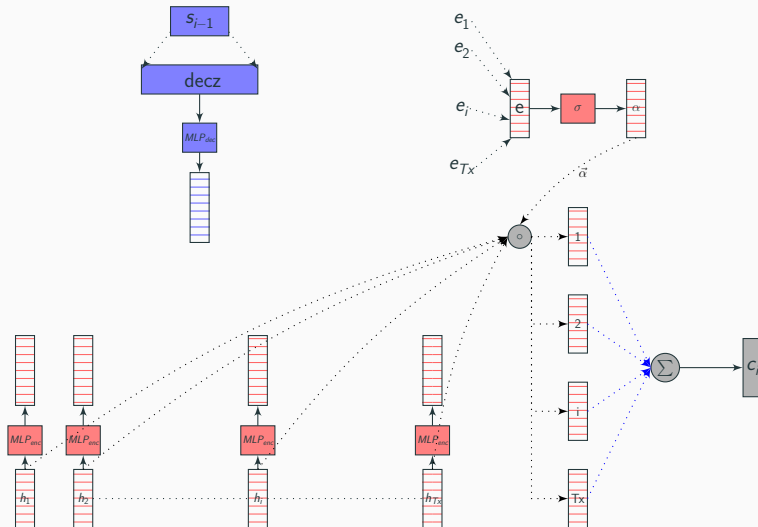




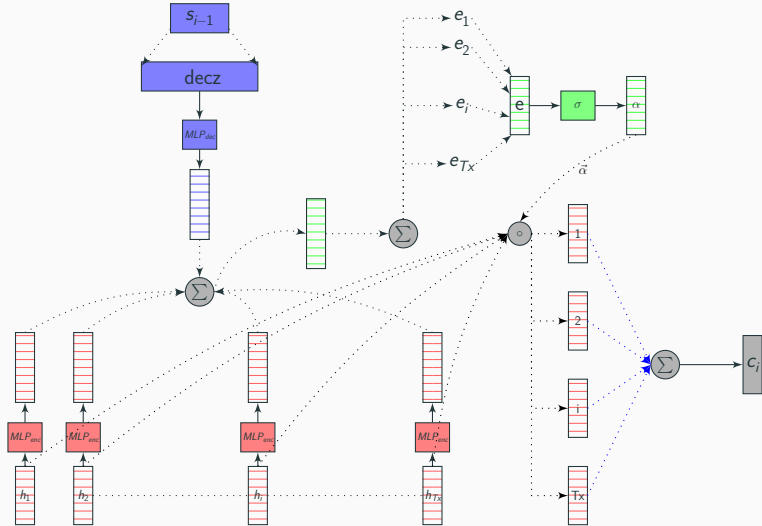




# Additive Attention

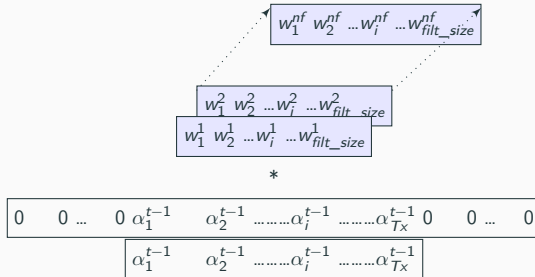


# Additive Attention - Full picture



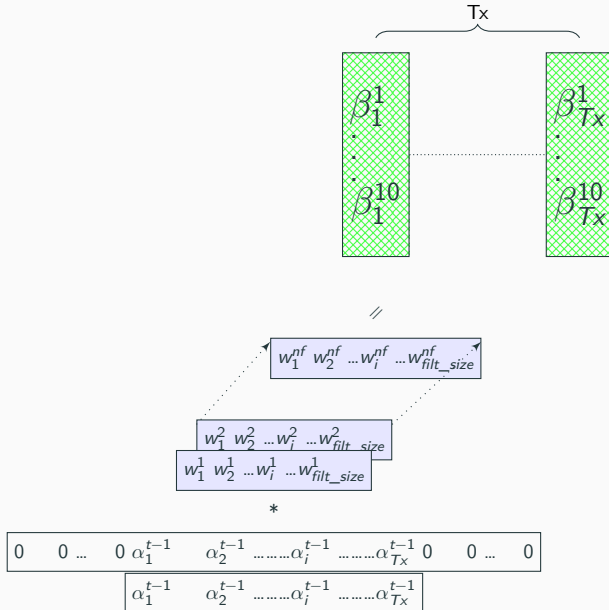
# Location Aware Attention

---

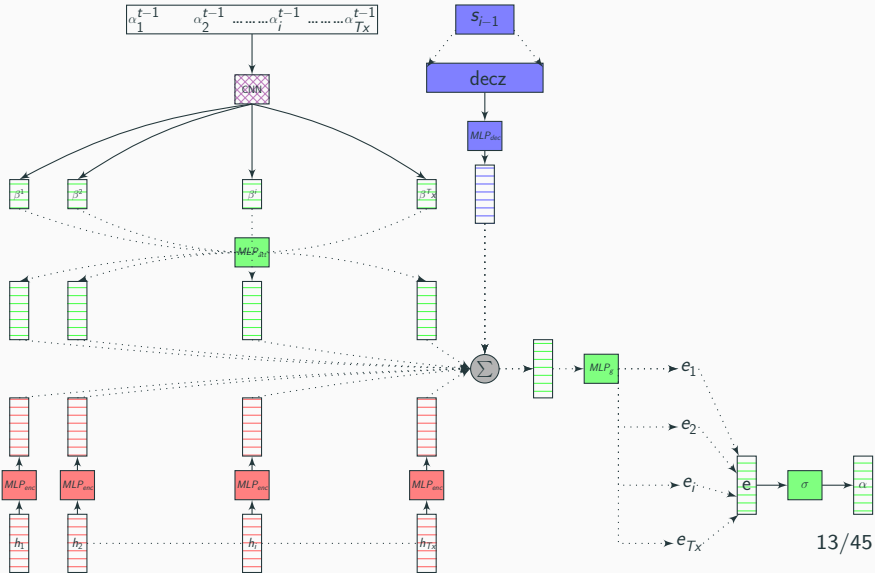


$$\begin{array}{c}
 \boxed{\beta_1^1 \quad \beta_2^1 \quad \dots \quad \beta_i^1 \quad \dots \quad \beta_{T_x}^1} \\
 \parallel \\
 \boxed{w_1^1 \quad w_2^1 \quad \dots \quad w_i^1 \quad \dots \quad w_{filt\_size}^1} \\
 * \\
 \boxed{0 \quad 0 \quad \dots \quad 0 \quad \alpha_1^{t-1} \quad \alpha_2^{t-1} \quad \dots \quad \alpha_i^{t-1} \quad \dots \quad \alpha_{T_x}^{t-1} \quad 0 \quad 0 \quad \dots \quad 0}
 \end{array}$$

$$\begin{array}{c}
 \boxed{\beta_1^2 \quad \beta_2^2 \quad \dots \quad \beta_i^2 \quad \dots \quad \beta_{T_x}^2} \\
 \parallel \\
 \boxed{w_1^2 \quad w_2^2 \quad \dots \quad w_i^2 \quad \dots \quad w_{filt\_size}^2} \\
 * \\
 \boxed{0 \quad 0 \quad \dots \quad 0 \quad \alpha_1^{t-1} \quad \alpha_2^{t-1} \quad \dots \quad \alpha_i^{t-1} \quad \dots \quad \alpha_{T_x}^{t-1} \quad 0 \quad 0 \quad \dots \quad 0}
 \end{array}$$



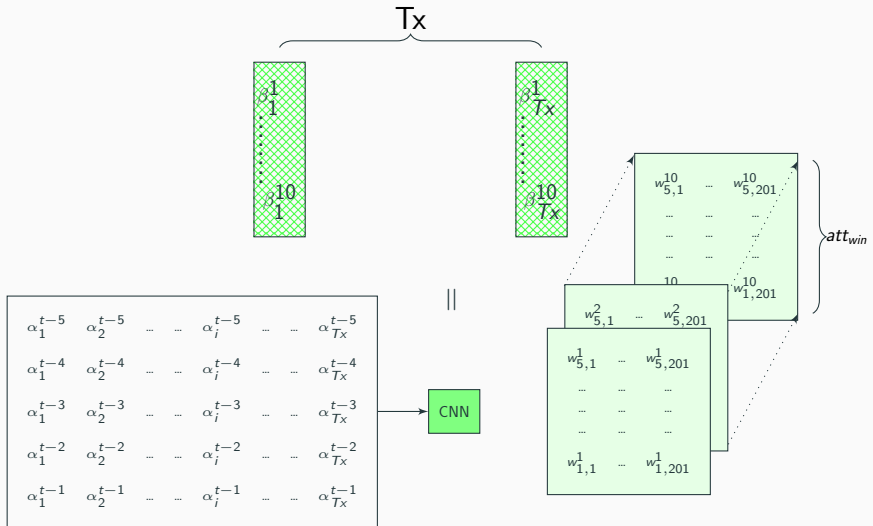
# Location Aware Attention - Full picture



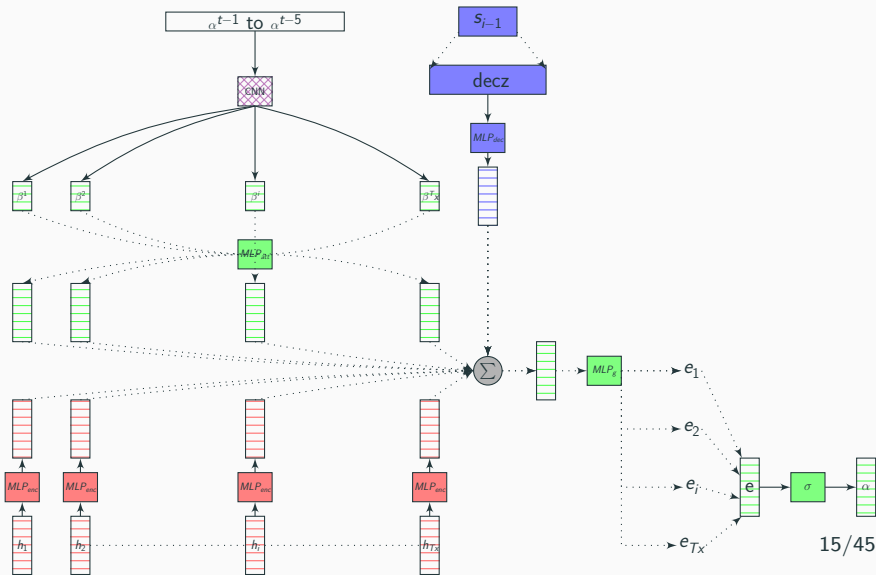


## 2D Location Aware Attention

---



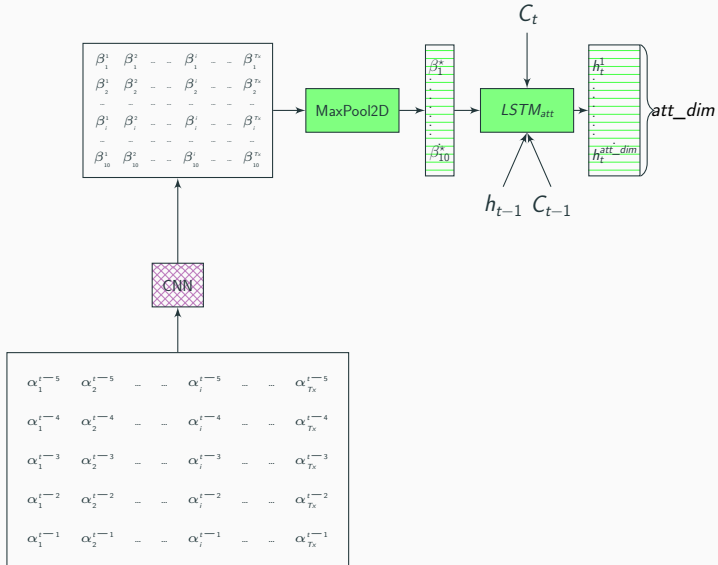
# 2D Location Aware Attention - Full picture



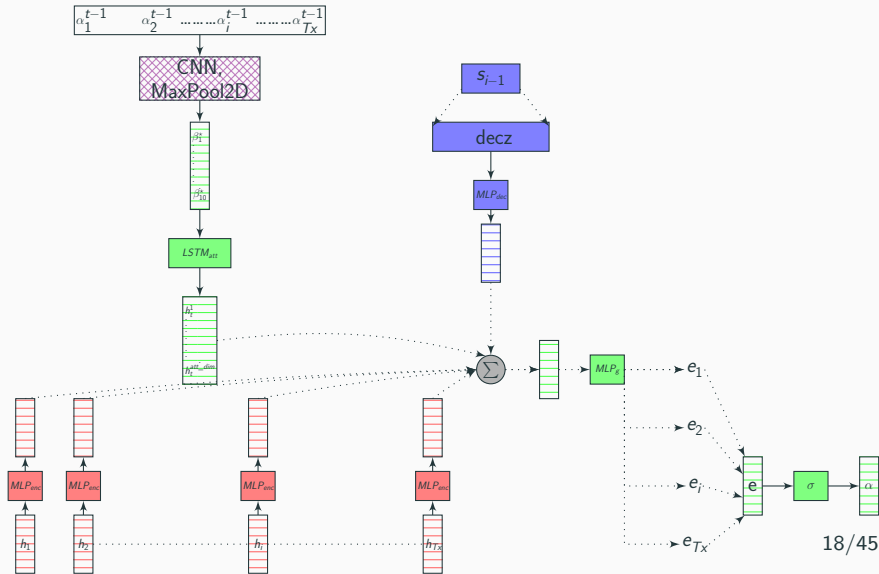
# Location Aware Recurrent Attention

---





# Location Aware Recurrent Attention - Full picture

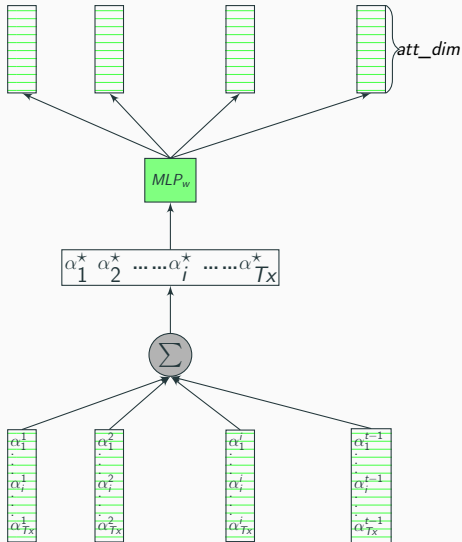


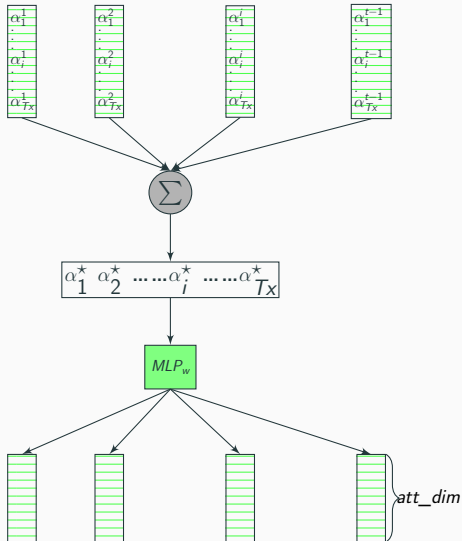
## Coverage mechanism **Attention**

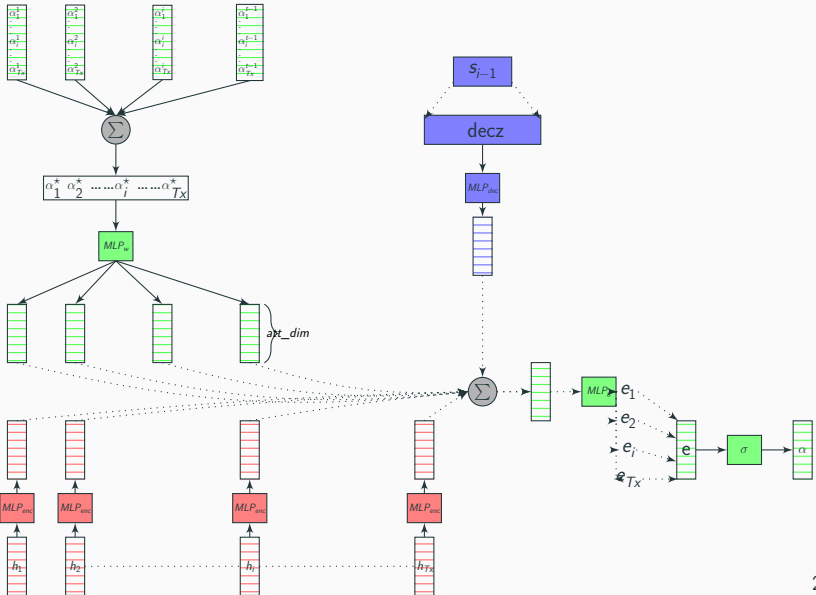
---

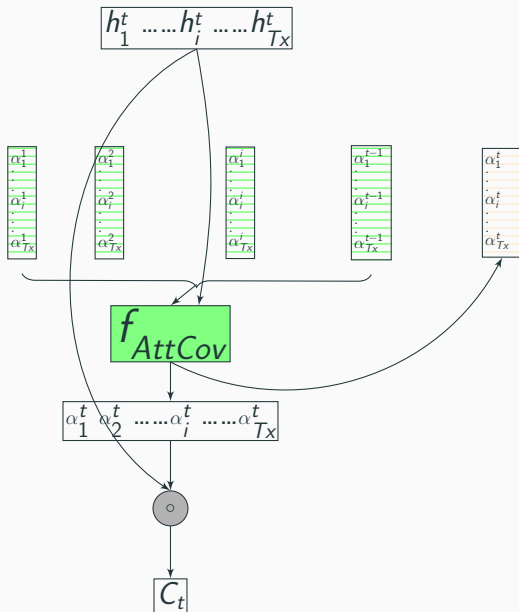


- Text summarization - Seq-to-Seq models
- Not reliable in producing factual details correctly
- Extend the standard seq-to-seq attention models
  - Hybrid pointer-generator network
  - Coverage









# Coverage mechanism location aware Attention

---

