

Optical Character Recognition (OCR)

- CV

Tesseract 3 (A*)

Binarizuje, naporcuje obrazok na kúsky a snaží sa z nich poskladať symboly

- CNN+RNN

Pre každý pixel povie pravdepodobnosť že je súčasťou textu, z oblastí z vysokou pravdepodobnosťou spraví obdĺžniky a tieto rozpoznáva rekurentnou neurónovou sieťou

- Holistic

Základuje obraz na príznakovú mapu a z nej generuje texty (transformer typu encoder-decoder)

Vstup

technical details are too complex to cover in the book itself.

In teaching our courses, we have found it useful for the students to attempt a number of small implementation projects, which often build on one another, in order to get them used to working with real-world images and the challenges that these present. The students are then asked to choose an individual topic for each of their small-group, final projects. (Sometimes these projects even turn into conference papers!) The exercises at the end of each chapter contain numerous suggestions for smaller mid-term projects, as well as more open-ended problems whose solutions are still active research topics. Wherever possible, I encourage students to try their algorithms on their own personal photographs, since this better motivates them, often leads to creative variants on the problems, and better acquaints them with the variety and complexity of real-world imagery.

In formulating and solving computer vision problems, I have often found it useful to draw inspiration from three high-level approaches:

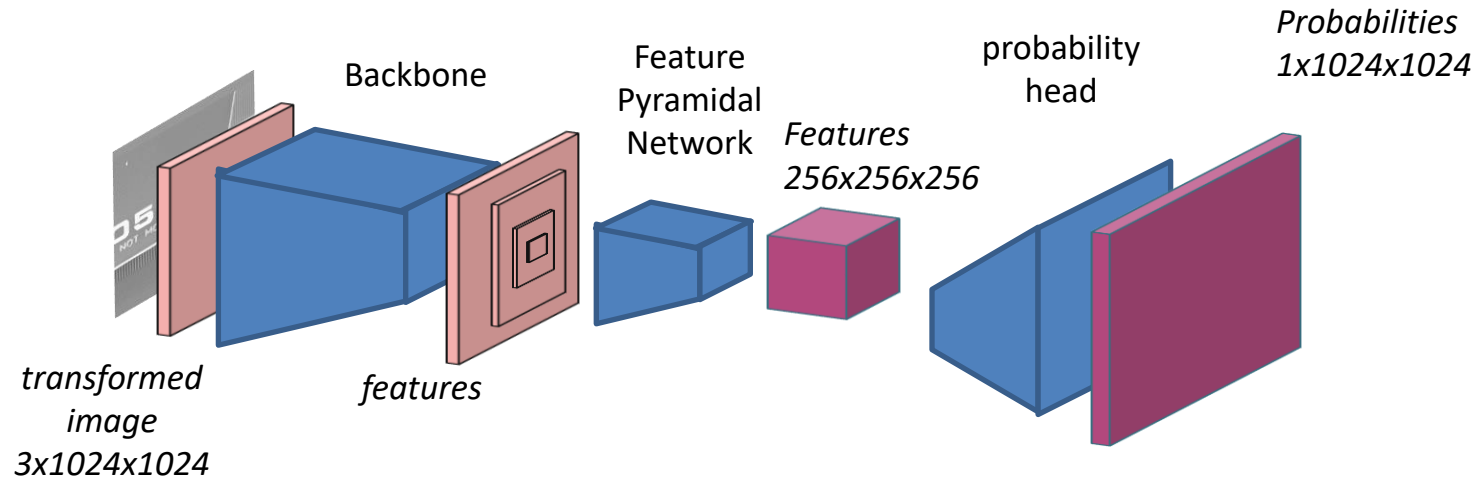
CNN+RNN

OCR tohto druhu sa skladá z dvoch hlbokých modelov:

- Detektor textu (konvolučná sieť)
- Rozpoznávač textu (dvojsmerná rekurentná sieť)

Výstup prvého je so vstupom druhého prepojený
algoritmami klasického počítačového videnia

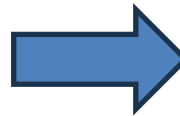
Detektor textu



technical details are too complex to cover in the book itself.

In teaching our courses, we have found it useful for the students to attempt a number of small implementation projects, which often build on one another, in order to get them used to working with real-world images and the challenges that these present. The students are then asked to choose an individual topic for each of their small-group, final projects. (Sometimes these projects even turn into conference papers!) The exercises at the end of each chapter contain numerous suggestions for smaller mid-term projects, as well as more open-ended problems whose solutions are still active research topics. Wherever possible, I encourage students to try their algorithms on their own personal photographs, since this better motivates them, often leads to creative variants on the problems, and better acquaints them with the variety and complexity of real-world imagery.

In formulating and solving computer vision problems, I have often found it useful to draw inspiration from three high-level approaches:



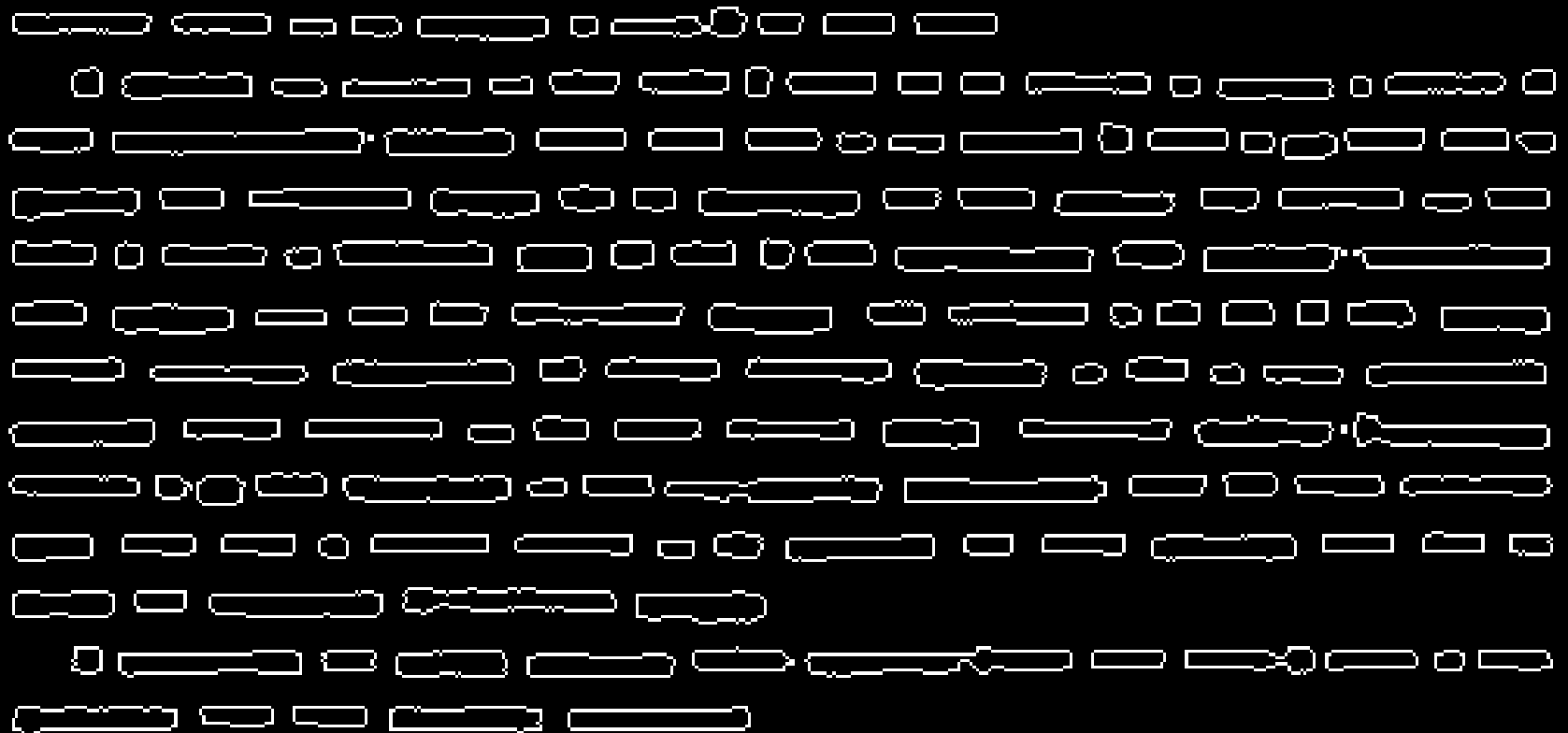
Výstup detektora

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Index**
 10. **Table of Contents**
 11. **Figure 1**
 12. **Figure 2**
 13. **Figure 3**
 14. **Figure 4**
 15. **Figure 5**
 16. **Figure 6**
 17. **Figure 7**
 18. **Figure 8**
 19. **Figure 9**
 20. **Figure 10**
 21. **Figure 11**
 22. **Figure 12**
 23. **Figure 13**
 24. **Figure 14**
 25. **Figure 15**
 26. **Figure 16**
 27. **Figure 17**
 28. **Figure 18**
 29. **Figure 19**
 30. **Figure 20**
 31. **Figure 21**
 32. **Figure 22**
 33. **Figure 23**
 34. **Figure 24**
 35. **Figure 25**
 36. **Figure 26**
 37. **Figure 27**
 38. **Figure 28**
 39. **Figure 29**
 40. **Figure 30**
 41. **Figure 31**
 42. **Figure 32**
 43. **Figure 33**
 44. **Figure 34**
 45. **Figure 35**
 46. **Figure 36**
 47. **Figure 37**
 48. **Figure 38**
 49. **Figure 39**
 50. **Figure 40**
 51. **Figure 41**
 52. **Figure 42**
 53. **Figure 43**
 54. **Figure 44**
 55. **Figure 45**
 56. **Figure 46**
 57. **Figure 47**
 58. **Figure 48**
 59. **Figure 49**
 60. **Figure 50**
 61. **Figure 51**
 62. **Figure 52**
 63. **Figure 53**
 64. **Figure 54**
 65. **Figure 55**
 66. **Figure 56**
 67. **Figure 57**
 68. **Figure 58**
 69. **Figure 59**
 70. **Figure 60**
 71. **Figure 61**
 72. **Figure 62**
 73. **Figure 63**
 74. **Figure 64**
 75. **Figure 65**
 76. **Figure 66**
 77. **Figure 67**
 78. **Figure 68**
 79. **Figure 69**
 80. **Figure 70**
 81. **Figure 71**
 82. **Figure 72**
 83. **Figure 73**
 84. **Figure 74**
 85. **Figure 75**
 86. **Figure 76**
 87. **Figure 77**
 88. **Figure 78**
 89. **Figure 79**
 90. **Figure 80**
 91. **Figure 81**
 92. **Figure 82**
 93. **Figure 83**
 94. **Figure 84**
 95. **Figure 85**
 96. **Figure 86**
 97. **Figure 87**
 98. **Figure 88**
 99. **Figure 89**
 100. **Figure 90**
 101. **Figure 91**
 102. **Figure 92**
 103. **Figure 93**
 104. **Figure 94**
 105. **Figure 95**
 106. **Figure 96**
 107. **Figure 97**
 108. **Figure 98**
 109. **Figure 99**
 110. **Figure 100**
 111. **Figure 101**
 112. **Figure 102**
 113. **Figure 103**
 114. **Figure 104**
 115. **Figure 105**
 116. **Figure 106**
 117. **Figure 107**
 118. **Figure 108**
 119. **Figure 109**
 120. **Figure 110**
 121. **Figure 111**
 122. **Figure 112**
 123. **Figure 113**
 124. **Figure 114**
 125. **Figure 115**
 126. **Figure 116**
 127. **Figure 117**
 128. **Figure 118**
 129. **Figure 119**
 130. **Figure 120**
 131. **Figure 121**
 132. **Figure 122**
 133. **Figure 123**
 134. **Figure 124**
 135. **Figure 125**
 136. **Figure 126**
 137. **Figure 127**
 138. **Figure 128**
 139. **Figure 129**
 140. **Figure 130**
 141. **Figure 131**
 142. **Figure 132**
 143. **Figure 133**
 144. **Figure 134**
 145. **Figure 135**
 146. **Figure 136**
 147. **Figure 137**
 148. **Figure 138**
 149. **Figure 139**
 150. **Figure 140**
 151. **Figure 141**
 152. **Figure 142**
 153. **Figure 143**
 154. **Figure 144**
 155. **Figure 145**
 156. **Figure 146**
 157. **Figure 147**
 158. **Figure 148**
 159. **Figure 149**
 160. **Figure 150**
 161. **Figure 151**
 162. **Figure 152**
 163. **Figure 153**
 164. **Figure 154**
 165. **Figure 155**
 166. **Figure 156**
 167. **Figure 157**
 168. **Figure 158**
 169. **Figure 159**
 170. **Figure 160**
 171. **Figure 161**
 172. **Figure 162**
 173. **Figure 163**
 174. **Figure 164**
 175. **Figure 165**
 176. **Figure 166**
 177. **Figure 167**
 178. **Figure 168**
 179. **Figure 169**
 180. **Figure 170**
 181. **Figure 171**
 182. **Figure 172**
 183. **Figure 173**
 184. **Figure 174**
 185. **Figure 175**
 186. **Figure 176**
 187. **Figure 177**
 188. **Figure 178**
 189. **Figure 179**
 190. **Figure 180**
 191. **Figure 181**
 192. **Figure 182**
 193. **Figure 183**
 194. **Figure 184**
 195. **Figure 185**
 196. **Figure 186**
 197. **Figure 187**
 198. **Figure 188**
 199. **Figure 189**
 200. **Figure 190**
 201. **Figure 191**
 202. **Figure 192**
 203. **Figure 193**
 204. **Figure 194**
 205. **Figure 195**
 206. **Figure 196**
 207. **Figure 197**
 208. **Figure 198**
 209. **Figure 199**
 210. **Figure 200**
 211. **Figure 201**
 212. **Figure 202**
 213. **Figure 203**
 214. **Figure 204**
 215. **Figure 205**
 216. **Figure 206**
 217. **Figure 207**
 218

It appears the Bureau has not failed to take the necessary steps to ensure a complete and accurate investigation. Indeed, the fact that the Bureau has not failed to take the necessary steps to ensure a complete and accurate investigation is a testament to the Bureau's commitment to the highest standards of law enforcement. The Bureau's commitment to the highest standards of law enforcement is a testament to the Bureau's commitment to the highest standards of law enforcement.

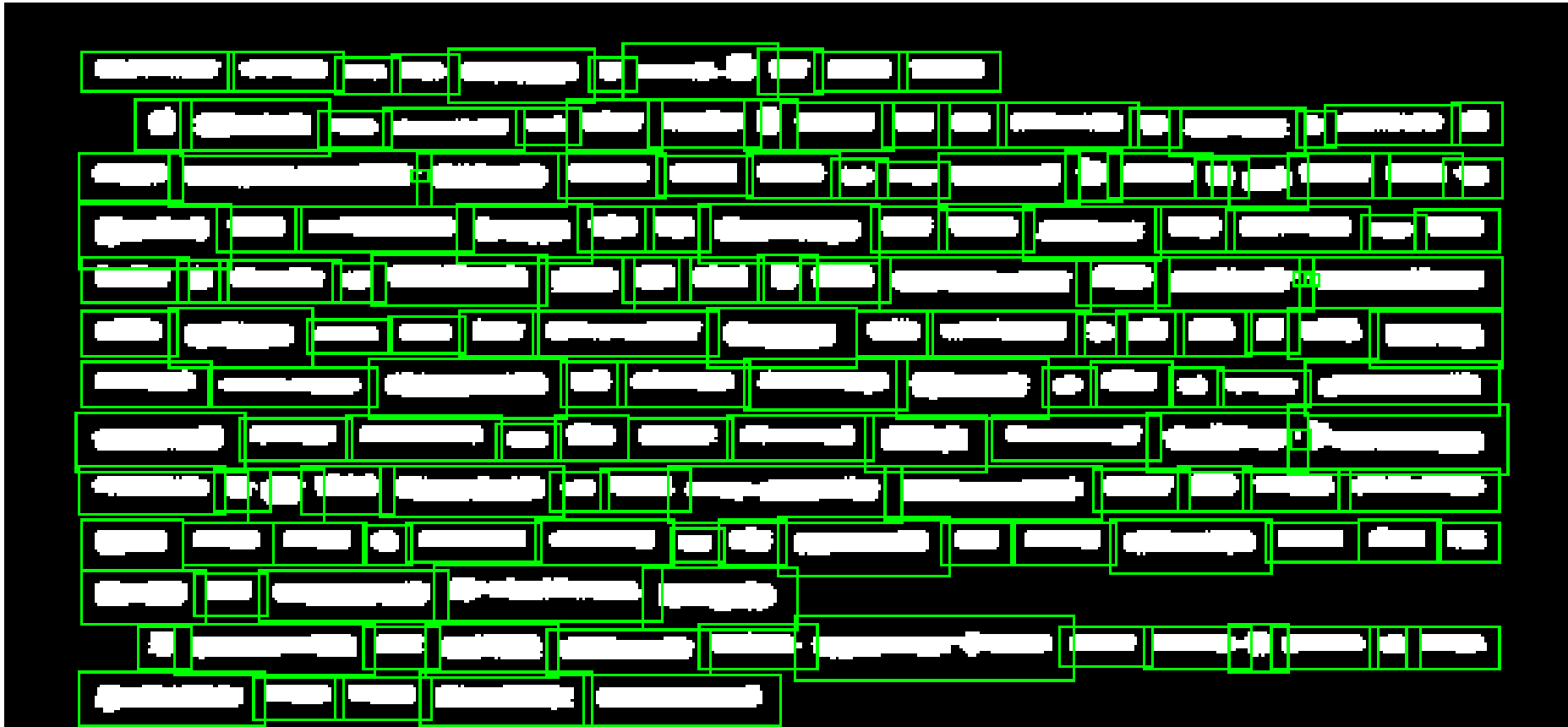
1. **Identify the main idea of the passage.**
 2. **Identify the supporting details.**
 3. **Identify the author's purpose.**
 4. **Identify the author's tone.**
 5. **Identify the author's point of view.**
 6. **Identify the author's bias.**
 7. **Identify the author's audience.**
 8. **Identify the author's style.**
 9. **Identify the author's structure.**
 10. **Identify the author's language.**

Kontúry



Kontúry môžeme aproximovať štyrmi alebo viacerými bodmi

Obdĺžniky alebo Polygóny



Pritom oblasti zodpovedajúco zväčšíme, aby sme zachytili celý text

Výsledok spracovania detekcie

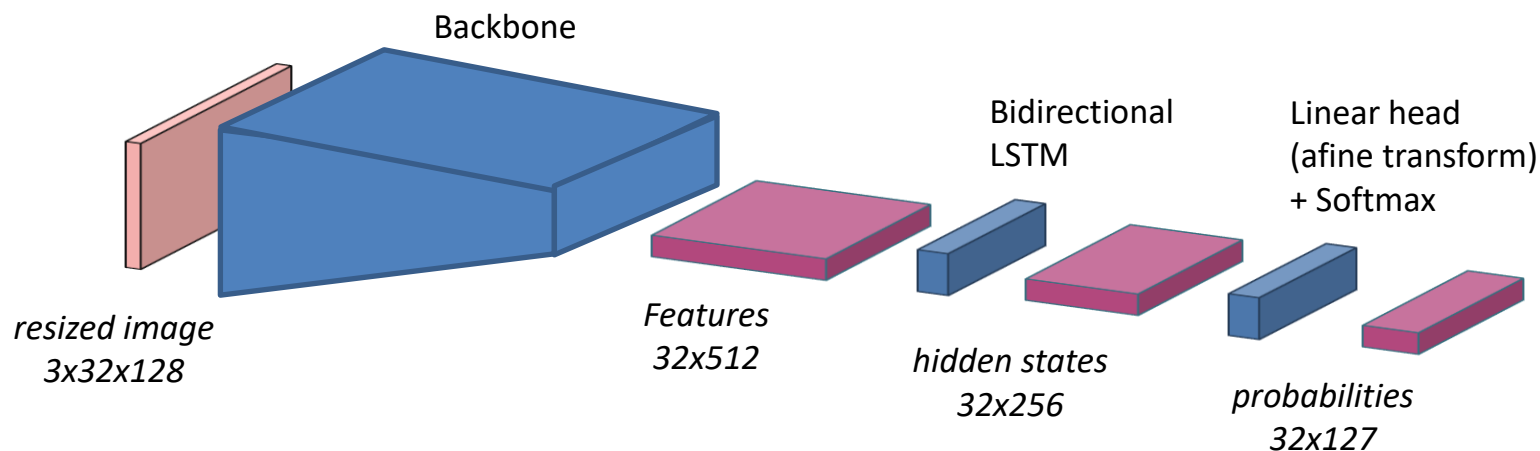
technical details are too complex to cover in the book itself.

In teaching our courses, we have found it useful for the students to attempt a number of small implementation projects, which often build on one another, in order to get them used to working with real-world images and the challenges that these present. The students are then asked to choose an individual topic for each of their small-group, final projects. (Sometimes these projects even turn into conference papers!) The exercises at the end of each chapter contain numerous suggestions for smaller mid-term projects, as well as more open-ended problems whose solutions are still active research topics. Wherever possible I encourage students to try their algorithms on their own personal photographs, since this better motivates them, often leads to creative variants on the problems, and better acquaints them with the variety and complexity of real-world imagery.

In formulating and solving computer vision problems, I have often found it useful to draw inspiration from three high-level approaches:

Každý obĺžnik či polygón vyjmene z textu a budeme rozpoznávať

Rozpoznávač



vocab (126 characters):

0123456789abcdefghijklmnopqrstuvwxyzABCD
EFGHIJKLMNOPQRSTUVWXYZ!"#\$%&'()*+,-
. / : ; < = > ? @ [\] ^ _ ` { | } ~ ° £ € ¥ ¢ ₪
à â é ê ë ï î ò ù û ü ç À Â É Ê Ë Ì Î Ï Ô Ù Ú Ü Ç

Rozpoznávanie

- Vyrežeme kúsok textu

approaches:

- Upravíme jeho veľkosť na 32x128

approaches:

- Tento obraz premeníme na mapu príznakov 32x512
- Tú dekodujeme cez LSTM (typ RNN) na skryté stavy 32x256
- Lineárnou hlavou ich premeníme na 32xN logitov, kde N je počet znakov + 1 (BLANK), aplikujeme softmax
- Cez CTC z pravdepodobností vyberieme symboly

Problém s rýchlosťou

- Dvojsmerná (bidirectional) RNN je schopná pracovať na texte rôznej dĺžky (je natáhovacia)
- Ale nevie pracovať na textoch rôznej dĺžky naraz v jednej dávke
- Preto DocTR radšej naseká texty tak, aby mali menej ako 32 symbolov a používame rovnako veľký vstup $3 \times 32 \times 128$
- (Ale Pytorch vie pustiť rôzne dĺžky naraz, akurát vyžaduje aby bola dávka utriedená podľa dĺžky zostupne takže treba urobiť správnu permutáciu)

Výstup

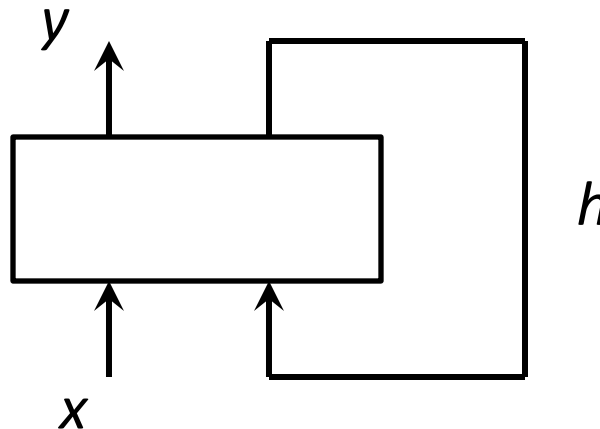
technical details are too complex to cover in the book itself.

In teaching our courses, we have found it useful for the students to attempt a number of small implementation projects, which often build on one another, in order to get them used to working with real-world images and the challenges that these present. The students are then asked to choose an individual topic for each of their small-group, final projects. (Sometimes these projects even turn into conference papers!) The exercises at the end of each chapter contain numerous suggestions for smaller mid-term projects, as well as more open-ended problems whose solutions are still active research topics. Wherever possible, I encourage students to try their algorithms on their own personal photographs, since this better motivates them, often leads to creative variants on the problems, and better acquaints them with the variety and complexity of real-world imagery.

In formulating and solving computer vision problems, I have often found it useful to draw inspiration from three high-level approaches:

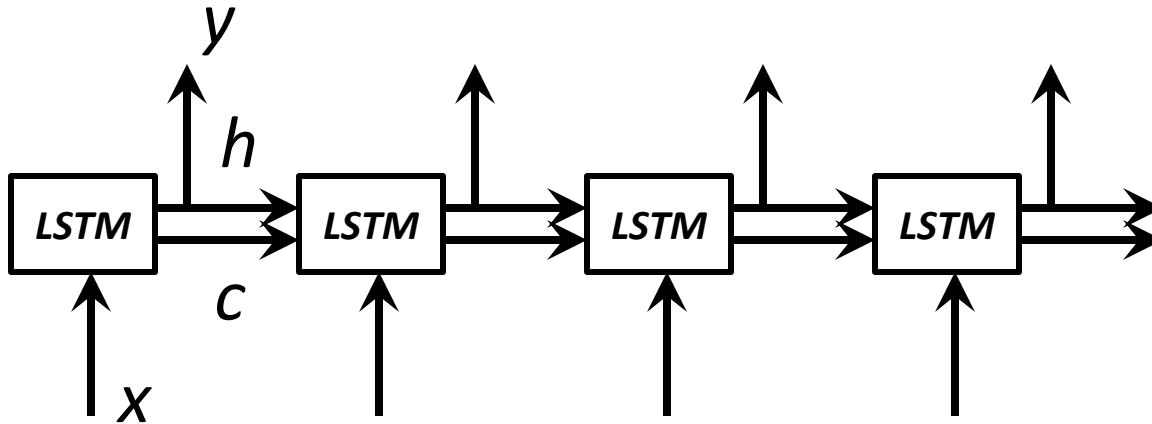
Rekurentná neurónová sieť

- Sieť inšpirovaná spracúvaním postupnosti dát v čase
- V každom okamihu dostáva nový vstup, dáva nový výstup a skrytý stav sa zapamätá a bude pridaný na vstup v ďalšom okamihu



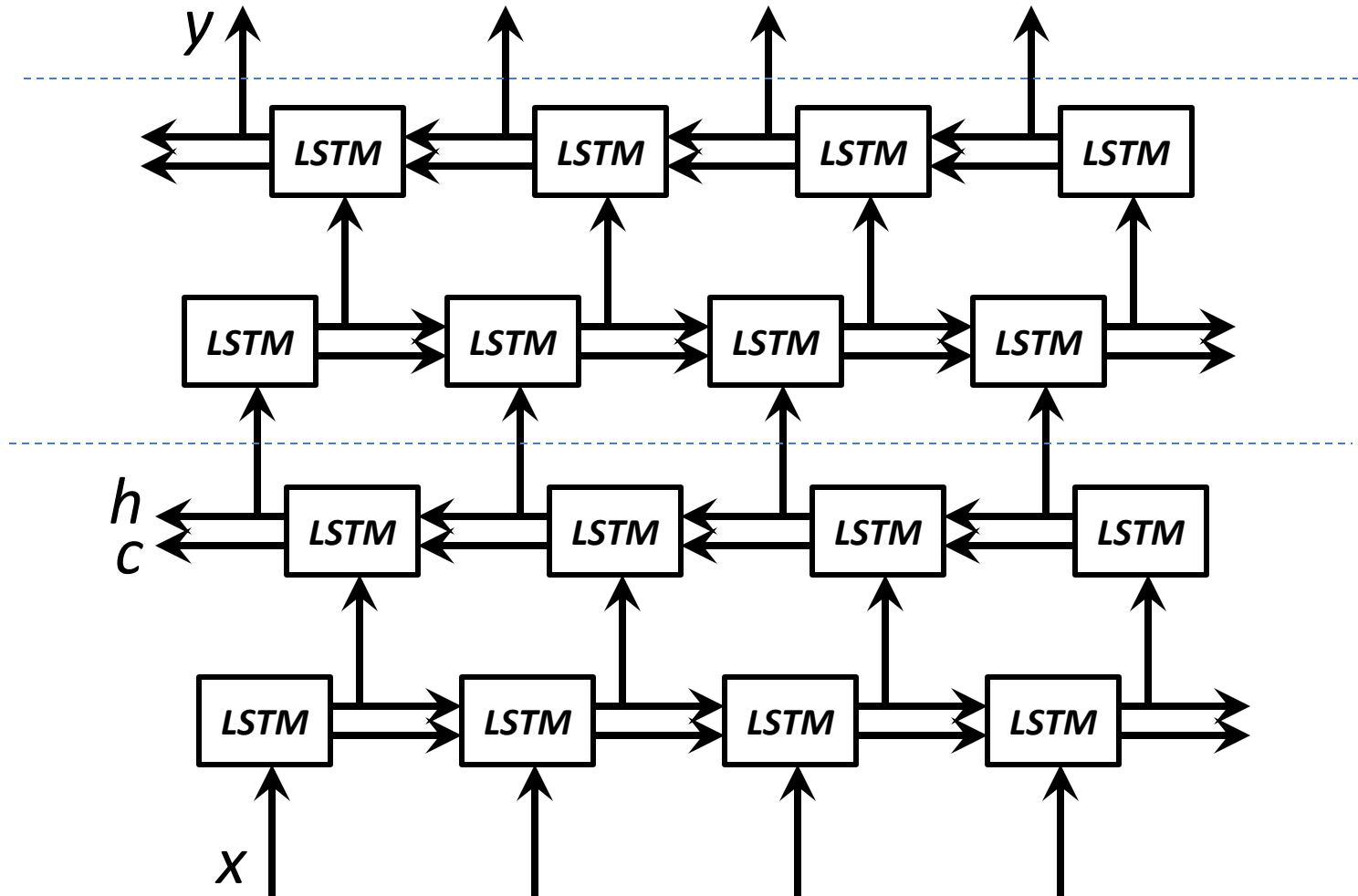
RNN

- Reálne sa nepúšťa na dáta postupne v čase, ale na konci naraz



- počet LSTM modulov je daný dĺžkou vstupu

Bidirectional RNN

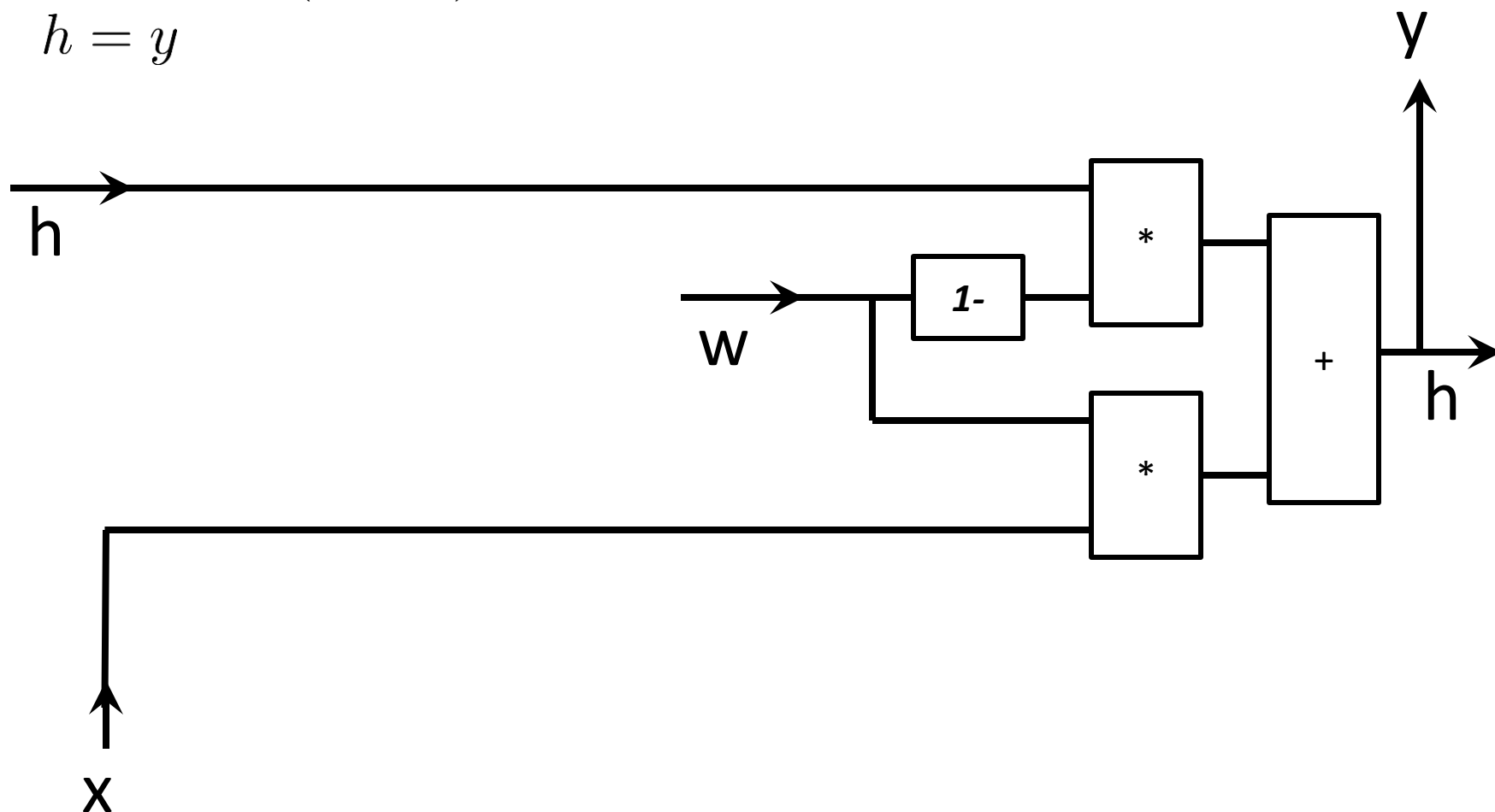


x: -1, -1, -1, -1, -1, -1, 1, 1, 1, 1, -1, 1, 1, 1, 1, 1, 1, -1, 1, 1, 1, 1, 1...

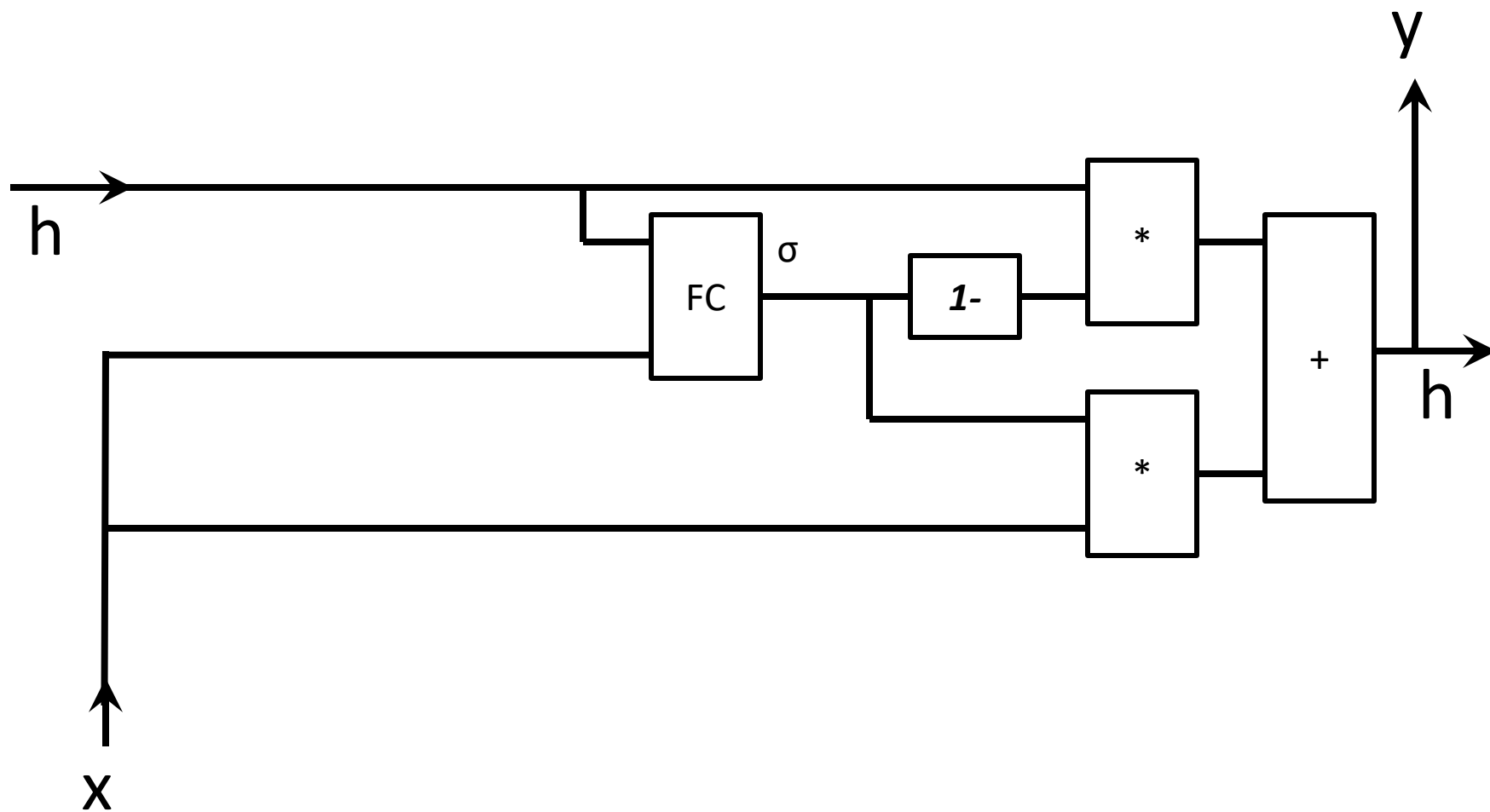
y: 0, -0.4, -0.7, -0.9, -1, -0.9, -0.7, -0.4, 0.1, 0.5, 0.9, 1, 1, 1, 1, ...

$$y = wx + (1 - w)h$$

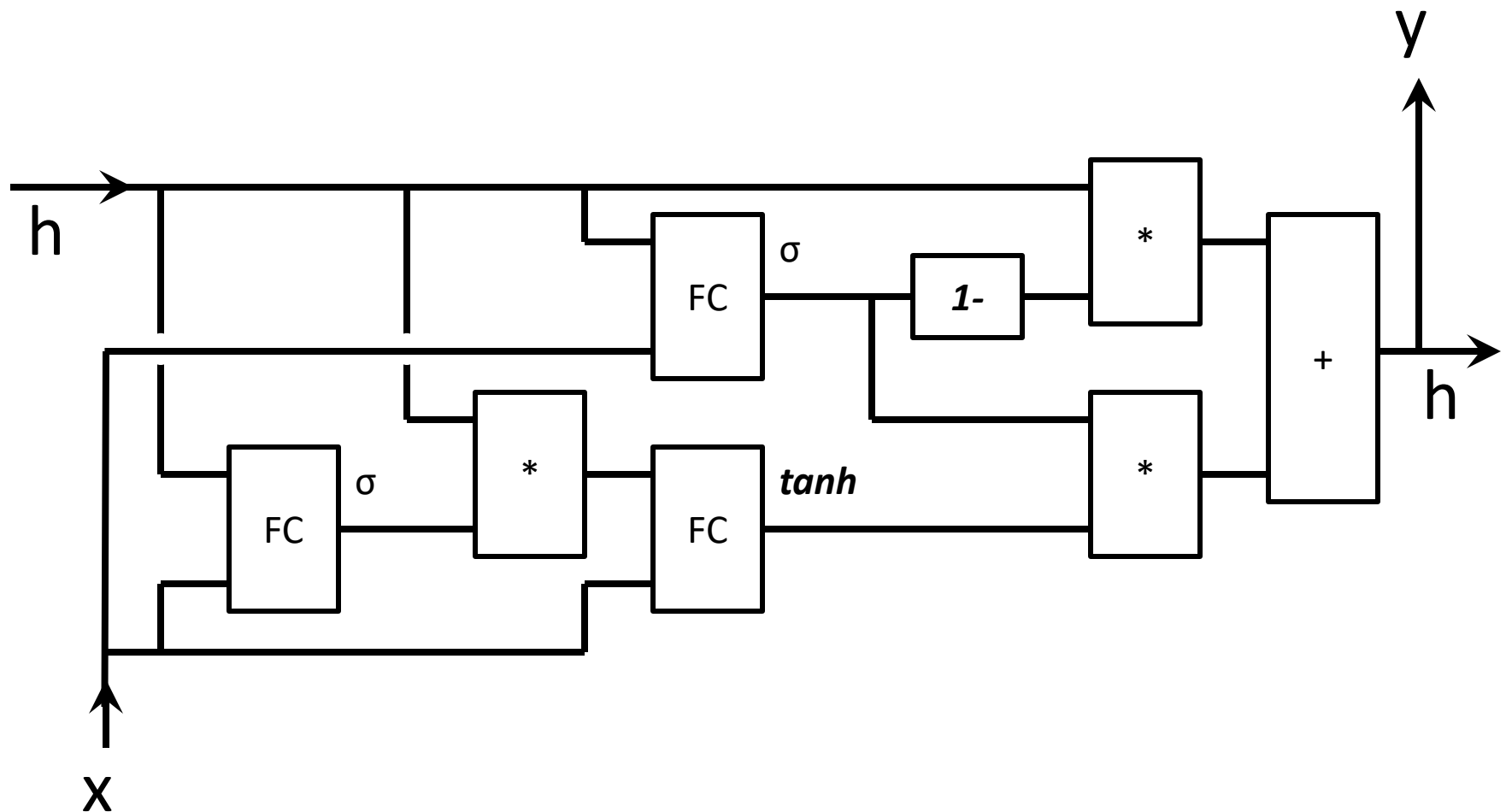
$$h = y$$



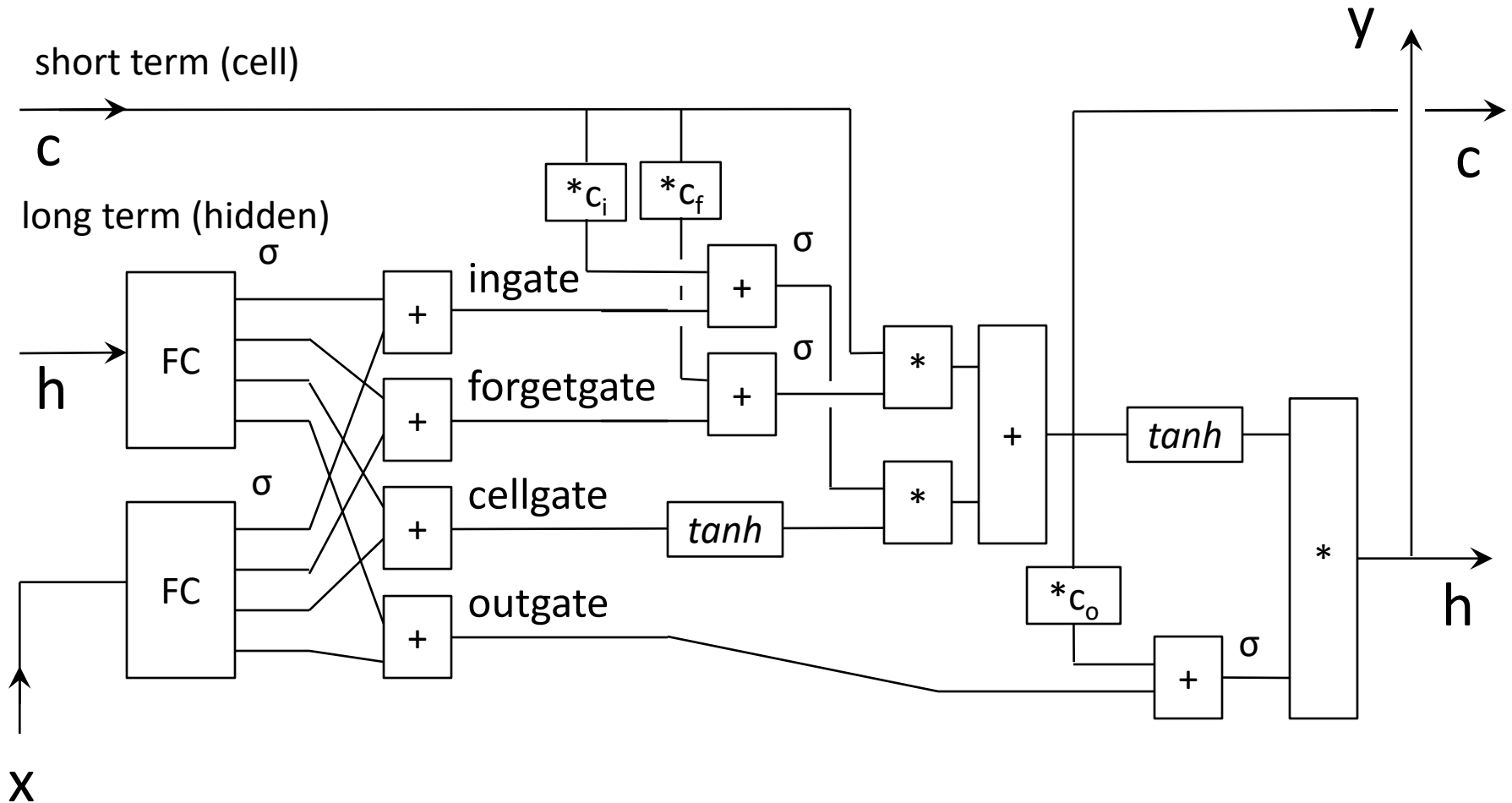
dimension of x, y and h can be > 1



Gated recurrent unit (GRU)



Long Short-Term Memory (LSTM)



Connectionist Temporal Classification (CTC)

- Chybová funkcia
- Počíta sumu pravdepodobností pre všetky možné zarovnaní
- Zarovnanie je cesta cez tenzor pravdepodobností, ktorá pod odstránení BLANKu dáva požadovanú sekvenciu; pravdepodobnosť cesty je súčinom pravdepodobností krokov
- Počíta sa efektívne cez dynamické programovanie

Connectionist Temporal Classification (CTC)

- Ako vedľajší produkt vie vypočítať najpravdepodobnejšiu cestu, takže ju je možné použiť aj na dekódovanie pri inferencii (beam search)
- Po natrénovaní však celkom dobre funguje aj greedy: pre každý výstup sa vezme najpravdepodobnejší symbol
- Rozdiel: beam nájde cestu s najväčším súčinom pravdepodobností, greedy s najväčším súčtom

Holistic OCR

Pracuje podobne ako chatbot:

- zakóduje obraz do mapy príznakov
- zakóduje textový prompt do sekvencie príznakov
- Zosype ich do jednej hromady, prípadne medzi nimi robí fúziu
- generuje v cykle text ako chatbot, pričom miesto zakódovanej otázky používa pri cross-attention danú hromadu príznakov

Holistic OCR

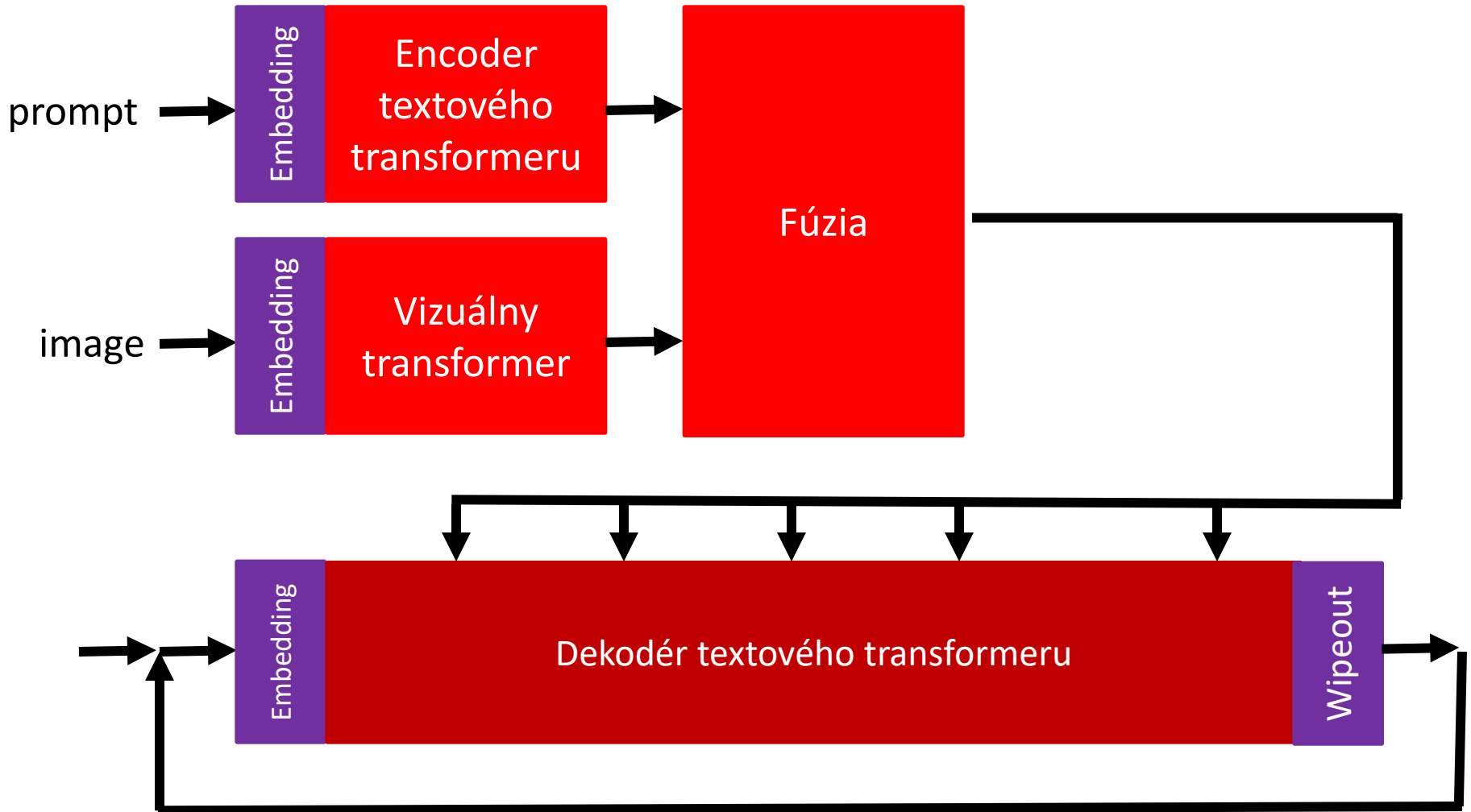
Výhody:

- Zohľadňuje jazyk
- Opravuje chyby
- Dosahuje výbornú presnosť

Nevýhody

- Vymýšľa si celé slová, ktoré na obraze nie sú
- Dokáže sa zacykliť !!!!

Holistic OCR



TrOCR: ukázky vstupu a výstupu

A is for apple

A IS FOR APPLE

A is for apple

A IS FOR APPLE

A is for apple

A IS FOR APPLE

A is for appl

A IS FOR APPLICA