- deep covnet → gradient vanishing when back prop

solve it thru {
- Residual Networks

F(x)

X (residual) → to solve the vanishing problem

F(x) + X
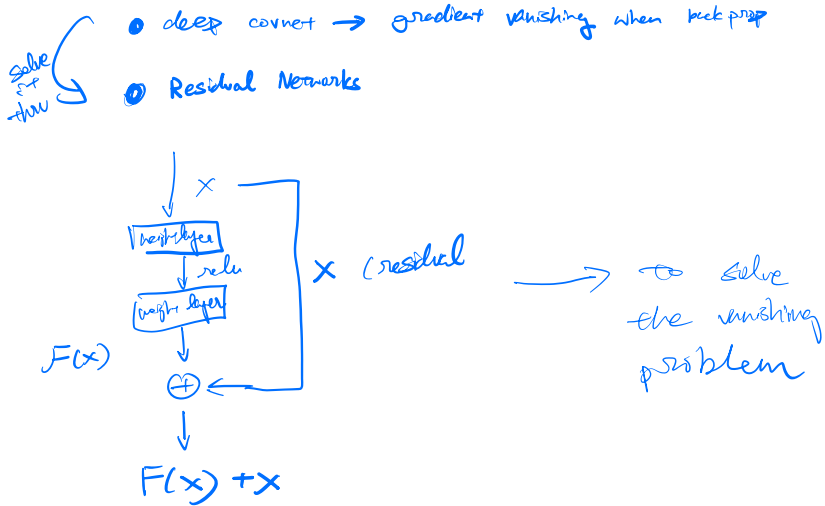
Natural Language Processing :

Glove

Word representation :
- Word Net
- one-hot vectors {
  - vector size fixed
  - sparsity
  - no natural notion of similarity.

word-coding matrix?

PPMI matrix → very sparse.

for sparsity : do SVD

$$\text{minimize} \quad ||M^{PPMI} - WV^T||_F^2$$
W. V

for binous element-wise summation

W×d   W×r   d×r

→ redundant info within the sparse matrix.

SVD : compression

Word2vec

on page 25
difference between two group of words,
if similar relations,
vector looks similar

problem: Bank of China v. river bank

↳ solution: contextual word embedding.

mission: text classification.

GRU $\rightarrow$ Simplified
of
LSTM