

VQA

By MedusaLafayetteDecorusScheisse

Recurrent Model :
LSTM(question) + Image Features(VGG) + MLP

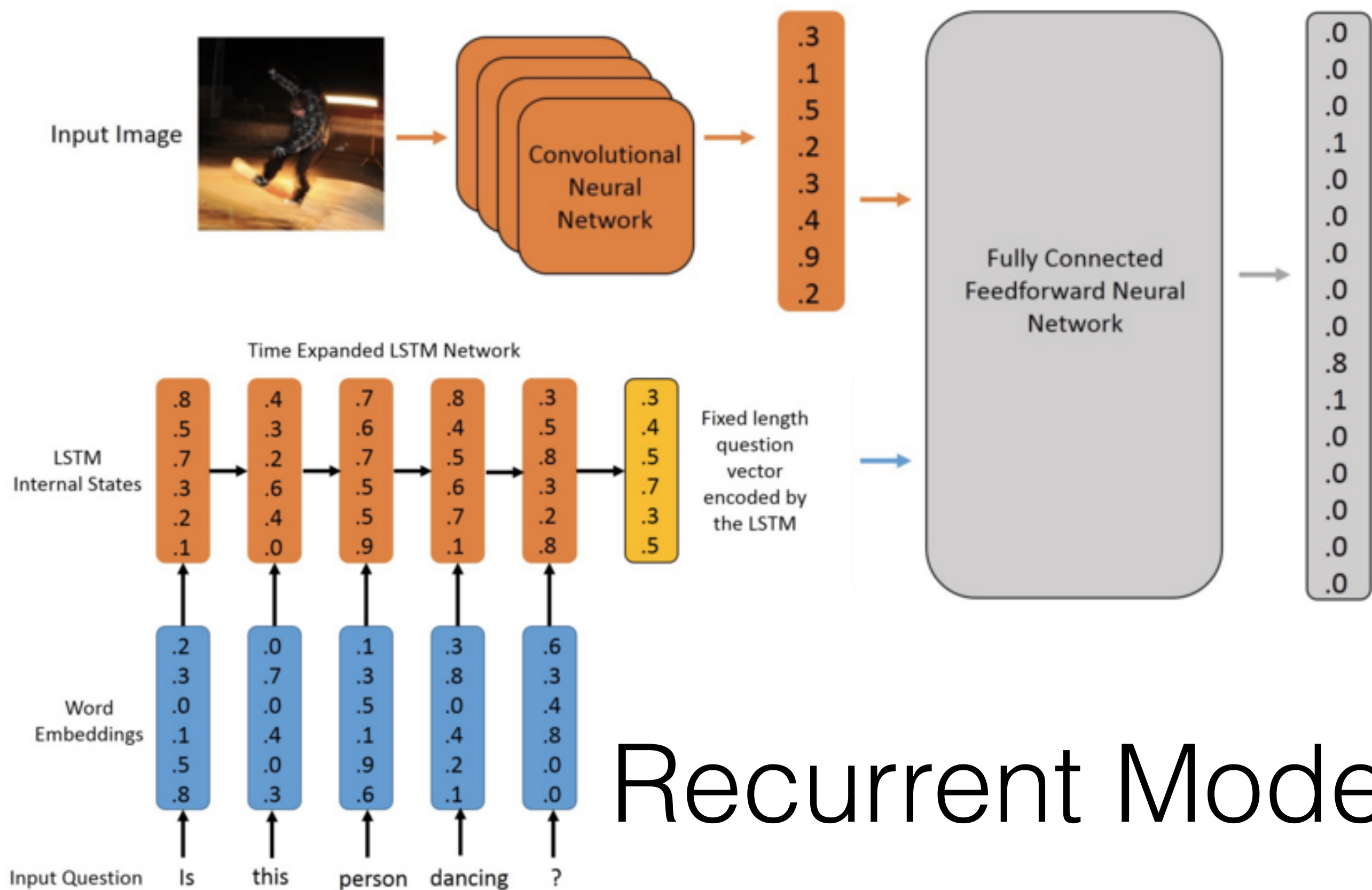


Use other material :
Image Features(Inception) + Caption



MemNN

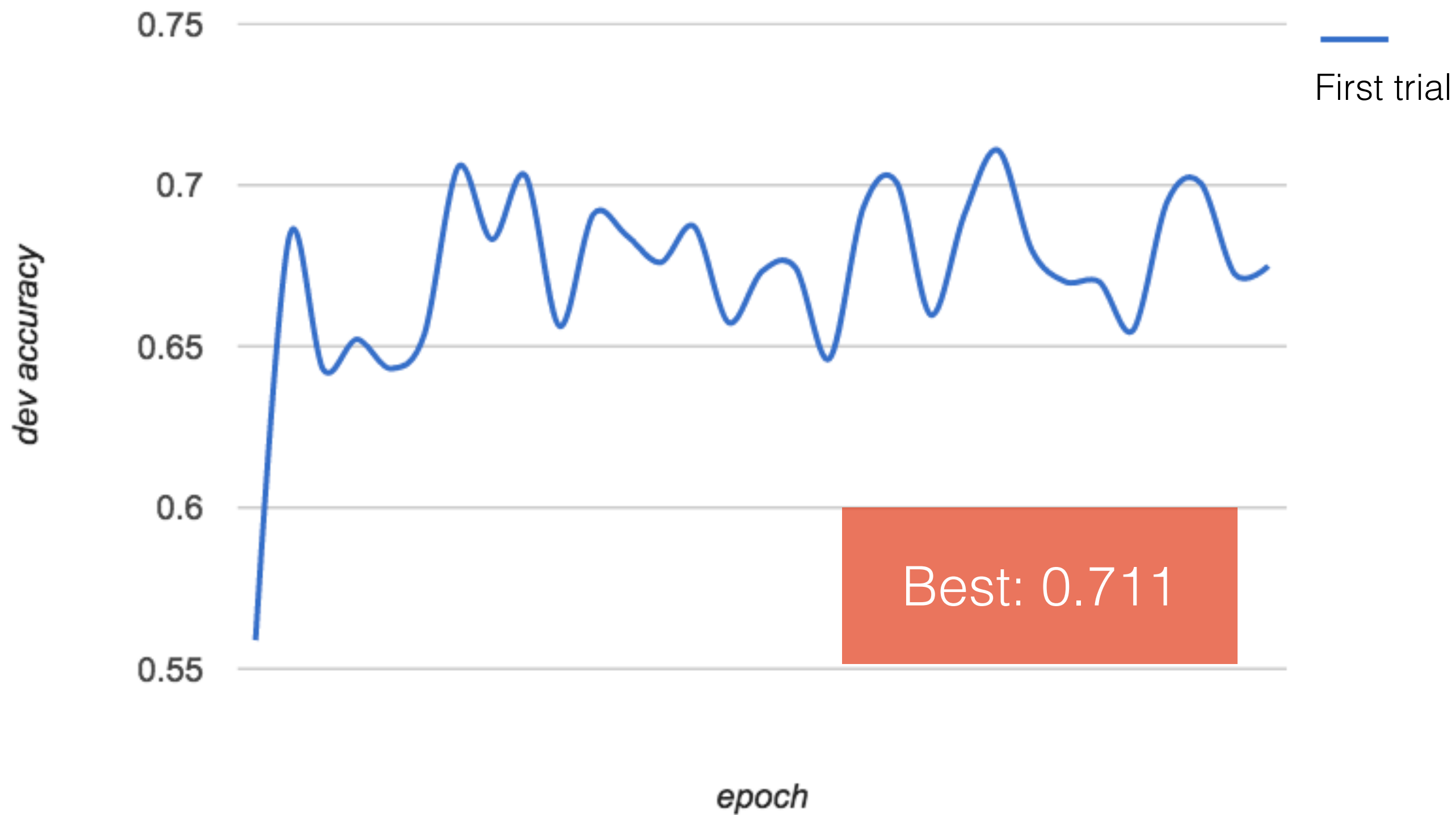
Recurrent Model :
LSTM(question) + Image Features(VGG) + MLP



Recurrent Model

First trial

	LSTM	MLP
layer	1	3
neurons	512	1024
dropout		0.5
learning rate		0.0001
activation		tanh



Tuning

LSTM

MLP

layer

1

3

neurons

512

1024

dropout

0.5

cost function

cosine

activation

relu

learning rate 4 trials : 0.00005 0.00002 0.00001
0.000005

Learning Rate	Best dev accuracy	Initialize dev accuracy
0.0005	0.721648	0.662208
0.0002	0.747483	0.662011
0.0001	0.744369	0.639061
0.00005	0.73607	0.597226

Activation Function Tuning

Activation function	Best dev accuracy	Initialize dev accuracy
Tanh	0.723025	0.633291
ReLu	0.721648	0.662208

Parsing	Best dev accuracy
Parsed by space tokens	0.72164
Parsed by Stanford parser	0.74856

Cost Function Tuning

Cost function	Best dev accuracy	Initialize dev accuracy
MSE	0.35402	0.26231
Cross entropy	0.74856	0.66047
Cosine similarity	0.74748	0.66201

Softmax?
Cross Entropy?

Output Activation function	Best dev accuracy	Initialize dev accuracy
Softmax	0.776414	0.696114
Linear Cosine similarity	0.78364	0.718473
Linear MSE	~0.53	-



MIT哥(誤?)想法的大腿, 一定要抱

Use other material :
Image Features(Inception) + caption

Google Inception

Inception

LSTM

MLP

layer

1

3

neurons

512

1024

dropout

0.3

cost function

cosine

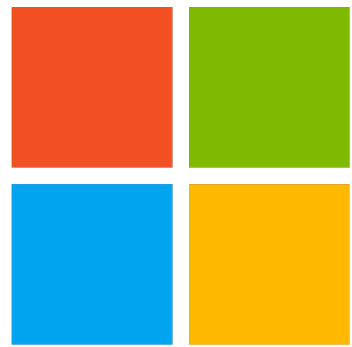
activation

relu

learning rate

0.0002

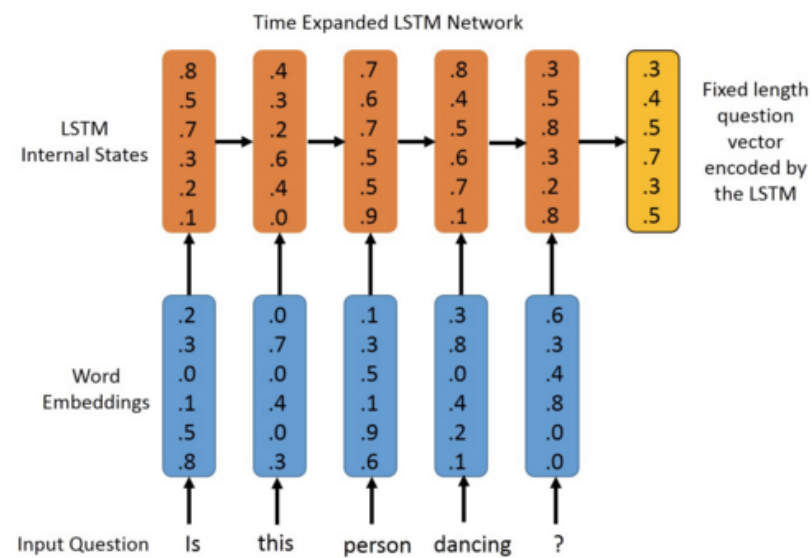
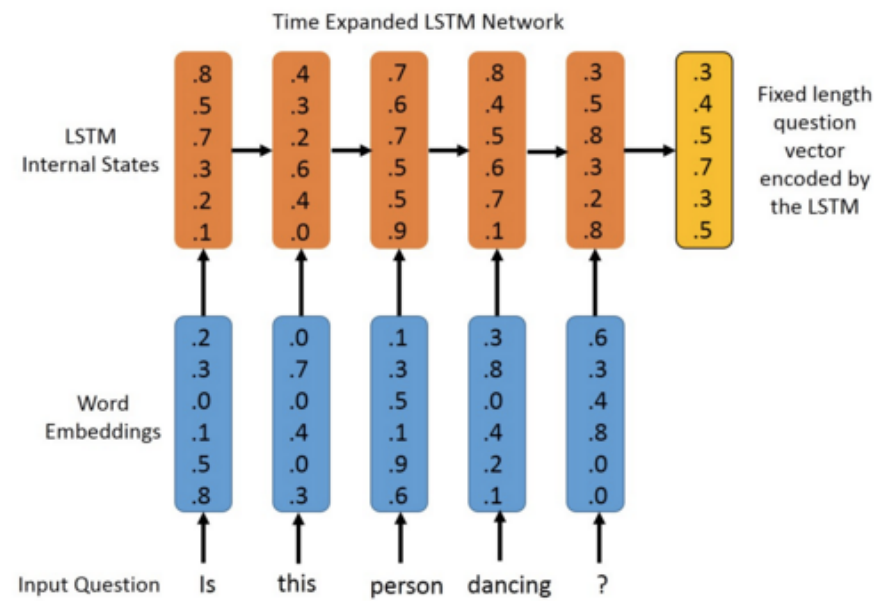
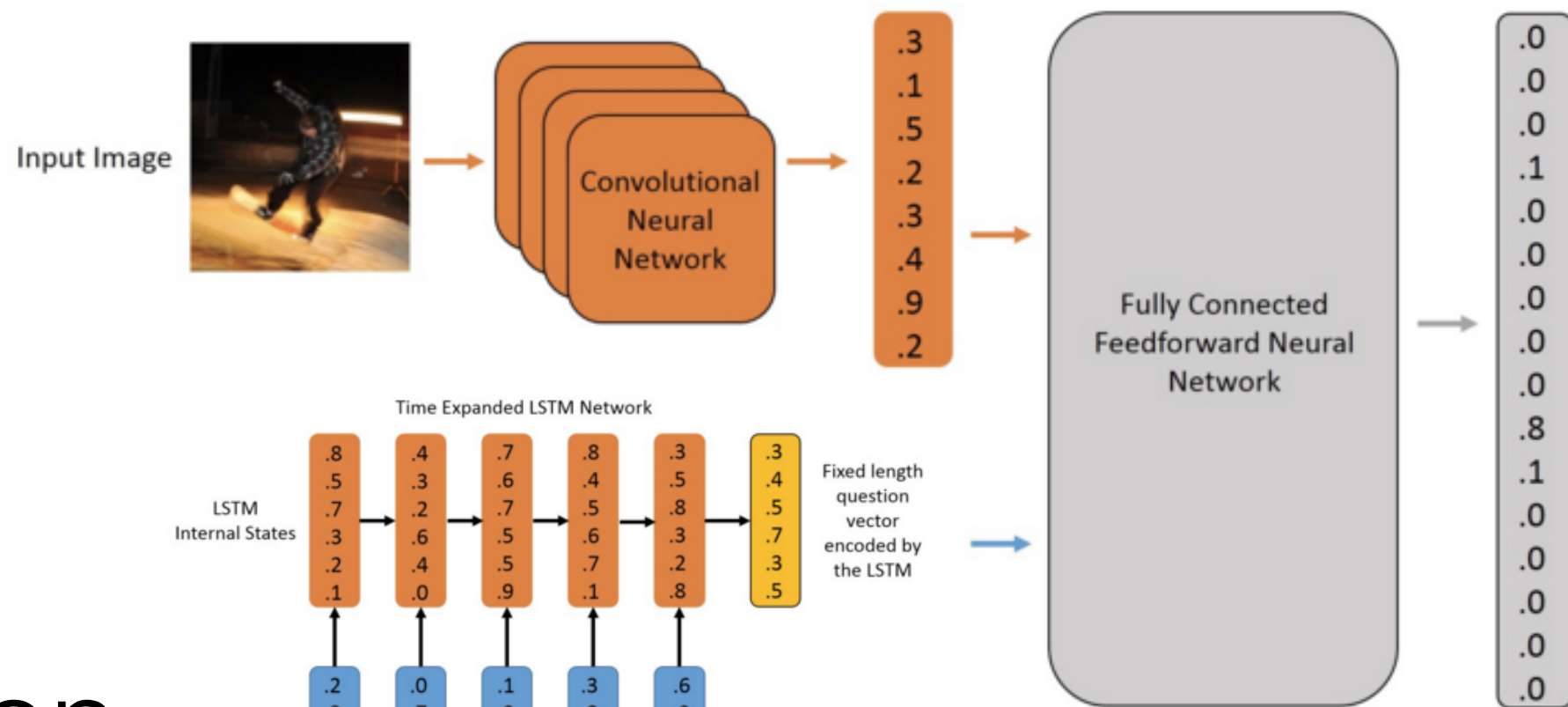
VGG Inception Ratio	Best dev accuracy
0.5 0.5	0.77384
0.25 0.75	0.77597
0.1 0.9	0.77446



Microsoft caption

question

caption

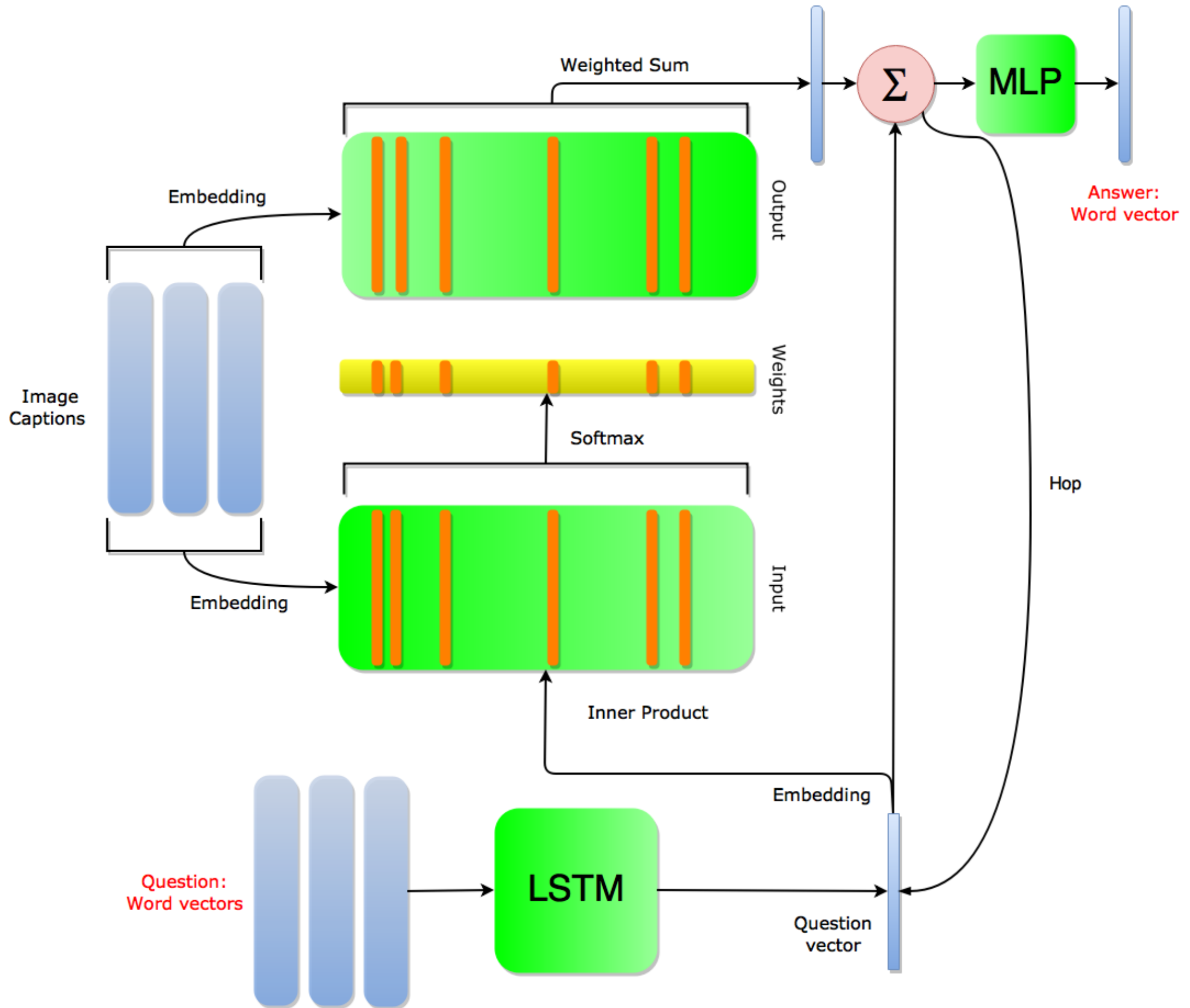


Material used	Best dev accuracy	Initialize dev accuracy
Inception*75+ VGG*0.25	0.77597	0.68119
Caption only	0.791	0.73374
Caption combine VGG and Inception	0.79832	0.72188

要站在Google, Microsoft巨人的肩膀上



MemNN



MemNN

hop
mem dim
memory size
layer
neurons
dropout
cost function
activation
learning rate

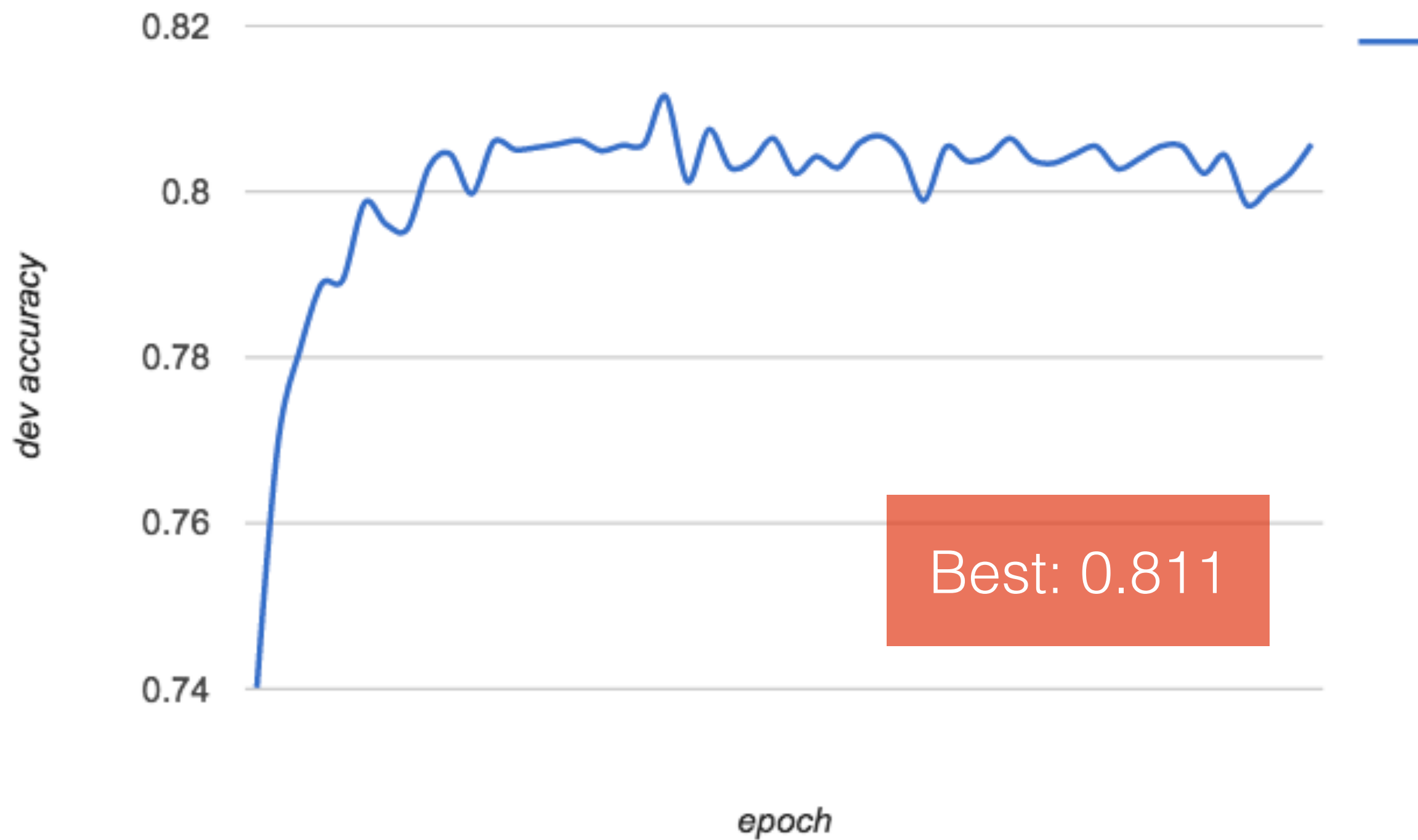
MemNN

2
50
125

MLP

1
1024
0
cosine
relu
0.0002

MemNN



MemNN Tuning

hop	Best dev accuracy	Initialize dev accuracy
1	0.77559	0.70552
2	0.81145	0.74015
3	0.80449	0.76087

MLP Layers	Best dev accuracy	Initialize dev accuracy
1	0.81145	0.74015
2	0.81363	0.75201
3	0.81295	0.75051

Mem dim	Best dev accuracy	Initialize dev accuracy
25	0.81145	0.74437
100	0.81213	0.76455
75	0.81607	0.73755

MemNN

hop
mem dim
memory size
layer
neurons
dropout
cost function
activation
learning rate

MemNN

2
75
125

MLP

2
1024
0
cosine
relu
0.0002



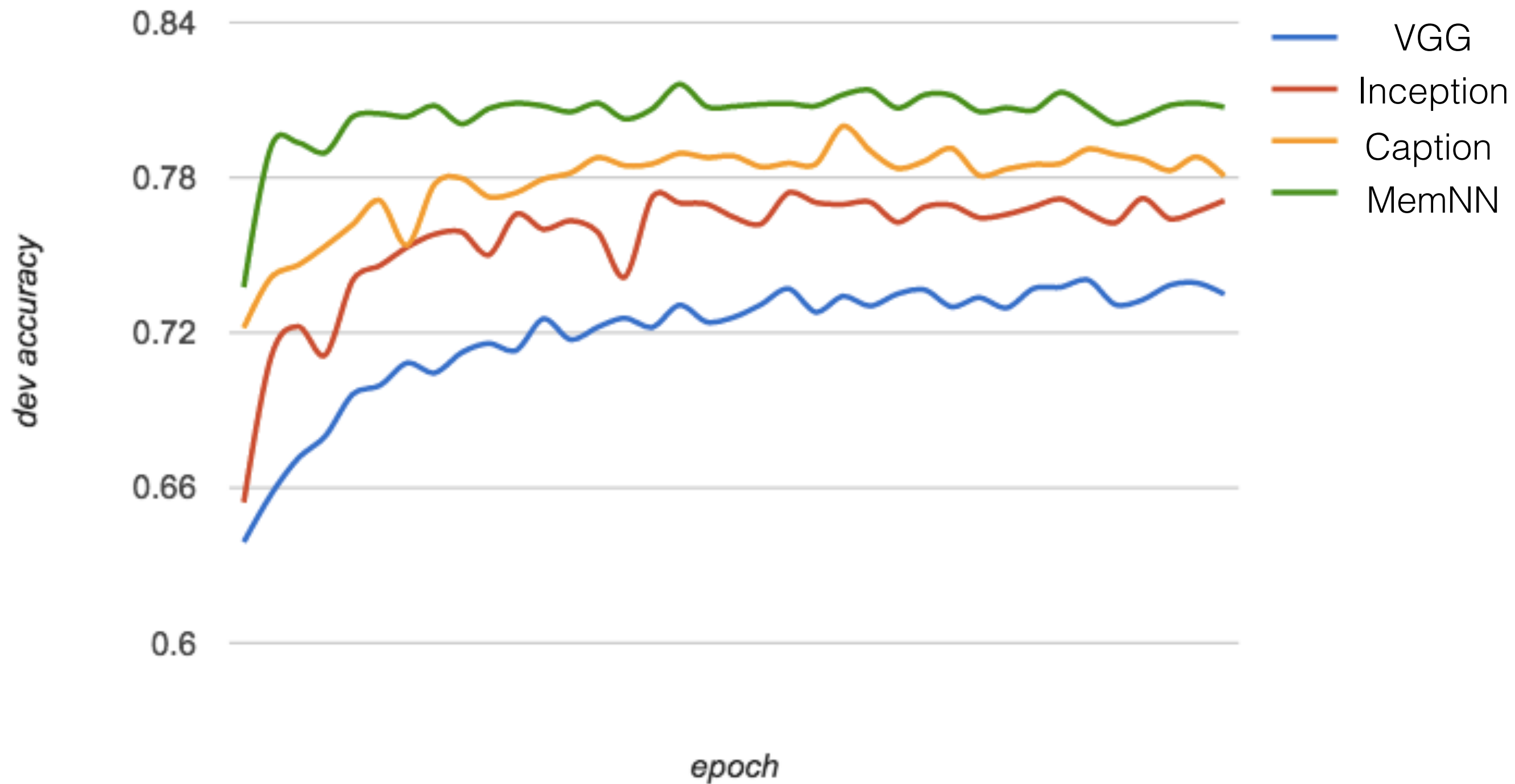
Handwritten signature in black ink.

Male God
我是男神

李宏毅老師上課的內容要實作

我是男神

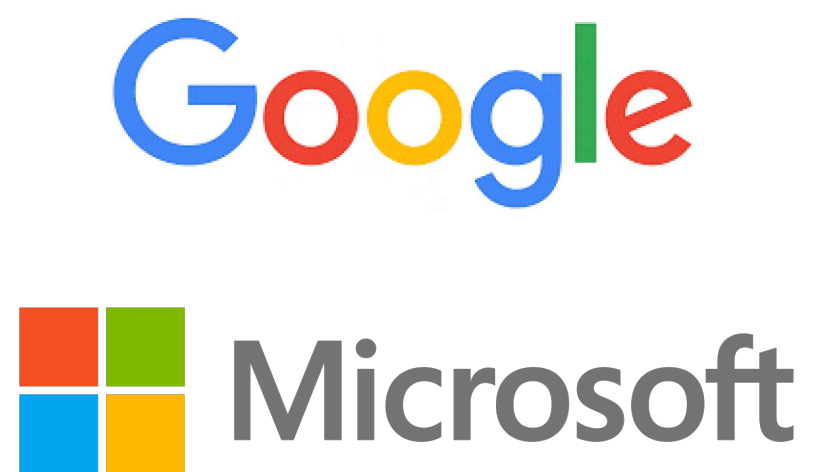
All Results



Conclusion



>



>

屌打



Thanks