Supplementary Material

Journal: Cognitive Computation

Paper Title: Dense-CaptionNet: A detailed sentence generation architecture for dense image captioning

Table 1: Qualitative Results – IAPR TC-12 dataset

Comparison with existing state-of-the-art techniques. The first row against each image illustrates the descriptions obtained (using 6,802 images of IAPR TC-12 dataset) on NeuralTalk (DeepVS) [1]. The 2nd-5th rows show the descriptions obtained by Show, Attend and Tell (Hard Attention) [2], SemStyle [3], Up-Down-Captioner [4] and the proposed Dense-CaptionNet, respectively.

DeepVS [1]	A large body of water with a boat in the background.	A man holding a surfboard on top of a beach.
Hard Attention [2]	A view of a large body of water.	A man standing in the middle of a river.
SemStyle [3]	I 'm going to be the beach.	The man who held a knife.
Up-Down-Captioner [4]	A large body of water on a beach.	A man holding a frisbee on a beach.
Dense-CaptionNet	**A body of water with blue water, white clouds in a blue sky in the background.**	**A man in white shirt and shorts standing on a beach holding onto a skateboard.**

DeepVS [1]	A large body of water with a bridge in the background.	A man riding a wave on top of a surfboard.
Hard Attention [2]	A view of a large body of water.	A person riding a snowboard down a snow covered slope.
SemStyle [3]	I 'll be coming up to the mountain.	A man who rode wave on top of the top of the surfboard.
Up-Down-Captioner [4]	A view of a city skyline with a clock tower.	A river in the water with a mountain in the water.
Dense-CaptionNet	**A scene to a white building, a blue sky in the city.**	**Trees on the shore, a landscape, a large mountain range in the distance.**
DeepVS [1]	A view of a mountain in the distance.	A man wearing a hat and a hat on a motorcycle.
Hard Attention [2]	A lone giraffe standing in the middle of a field.	A man wearing a helmet and a helmet on a motorcycle.
SemStyle [3]	I 'll be coming up to the mountain.	A man who wore a helmet.
Up-Down-Captioner [4]	A small tree next to a tree in a forest.	A close up of a person talking on a cell phone.
Dense-CaptionNet	**A small green tree with a tall green tree in the middle of an area.**	**A man on a motorcycle with a red shirt and a pair of shoes on the street.**

DeepVS [1]	A bedroom with a bed and a table.	A view of a lake with a bridge and a river.
Hard Attention [2]	A bedroom with a bed and a bed.	A large body of water with a boat in the background.
SemStyle [3]	I climb into bed and then i stormed into the room.	" i 'll be coming up to the mountain .
Up-Down-Captioner [4]	A bed with a bed and a bed in it.	A boat in the water near the water.
Dense-CaptionNet	**A bed with a comforter and a wooden headboard.**	**A boat with blue white water, a large body of water and a large rock in the water.**

Table 2: Qualitative Results on MSCOCO

Comparison with state-of-the-art techniques. The first row against each image shows MSCOCO descriptions. The 2nd- 6th rows show the descriptions obtained by NeuralTalk (DeepVS) [1], Show, Attend and Tell (Hard Attention) [2], SemStyle [3], Up-Down-Captioner [4] and the proposed Dense-CaptionNet, respectively.

MSCOCO Descriptions	- A home computer and mouse on a desk. - A desk that has a computer and other various items on it. - A movie is playing on the computer monitor. - A computer, key board, cell phones and other electronic gadgets are on the table. - A computer desk topped with a desktop computer monitor and keyboard.	- This is a man riding a red crotch rocket. - A man is riding a motorcycle on a road. - A man that is riding around on a motorcycle. - A motorcycle rider on a red silver and black motorcycle. - A man riding on the back of a red motorcycle.
DeepVS [1]	a desk with a laptop and a monitor	a man riding a motorcycle down a street
Hard Attention [2]	a laptop computer sitting on top of a desk.	a man riding a motorcycle down a street.
SemStyle [3]	i sat up at the computer desk and logged onto the computer .	the man rode his motorcycle on the road .
Up-Down-Captioner [4]	A computer desk with a computer and a keyboard.	A person riding a red motorcycle on the street.
Dense-CaptionNet	a desk with a laptop, a keyboard, a mouse and a monitor.	a man wearing a helmet is riding a motorcycle
MSCOCO Descriptions	- Small bird perched on a branch in a wooden area. - A small bird perched on top of a tree branch. - A small bird sitting on a small branch in	- A man is playing video games on a screen. - A man prepares to swing while playing a video game. - An person relishing it on a

	a tree. - A small bird perched on a branch of tree. - A bird sitting in a tree on a branch.	good daytime. - A person is playing wii on the tv in the room. - The boy is holding a remote as he plays a game
DeepVS [1]	a bird sitting on a branch in a tree	a man playing a video game in a living room
Hard Attention [2]	a bird sitting on top of a tree branch.	a man is playing a video game in a living room
SemStyle [3]	i was sitting at the bird 's sitting on the tree branch .	" the man who is playing games , " she said , " a game
Up-Down-Captioner [4]	A small bird perched on a tree branch.	A man is playing a video game on a television.
Dense-CaptionNet	A bird sitting on a tree branch with leaves on it	A man standing in front of a screen tv holding a game controller
MSCOCO Descriptions	- A computer sitting on a see through desk - A desktop computer sitting on top of a glass table - A home computer on a desk in a room - Many electronic devices and a tangle of wires are on and around the desk - A home office with a chrome and glass desk and computer equipment	- A computer monitor sitting on top of a computer desk. - A metal and glass computer desk by a window. - A modern glass and metal desk with multiple computers on it. - A computer monitor is on top of a glass desk. - A computer atop a glass table near a window.
DeepVS [1]	a desk with a computer monitor and keyboard	a desk with a computer monitor and keyboard
Hard Attention [2]	a desk with a computer and a monitor.	a desk with a computer and a monitor.
SemStyle [3]	i sat up at the computer desk.	i sat up at the office desk and logged onto the computer .
Up-Down-Captioner [4]	A desk with a computer and mouse on it.	A computer desk with a computer monitor and a monitor.
Dense-CaptionNet	A desk with a laptop a keyboard a mouse and a monitor	A desk with a laptop a keyboard a mouse and a monitor

References:

[1] A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generatingimage descriptions, IEEE Transactions on Pattern Analysis and MachineIntelligence 39 (4) (2017) 664–676.doi:10.1109/TPAMI.2016.2598339

[2] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel,Y. Bengio, Show, attend and tell: Neural image caption generation withvisual attention, in: International Conference on Machine Learning, 2015,pp. 2048–2057.

[3] A. Mathews, L. Xie, X. He, Semstyle: Learning to generate stylised imagecaptions using unaligned text, in: Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition, 2018, pp. 8591–8600.

[4] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, L. Zhang,Bottom-up and top-down attention for image captioning and visual ques-tion answering, in: CVPR, Vol. 3, 2018, p. 6