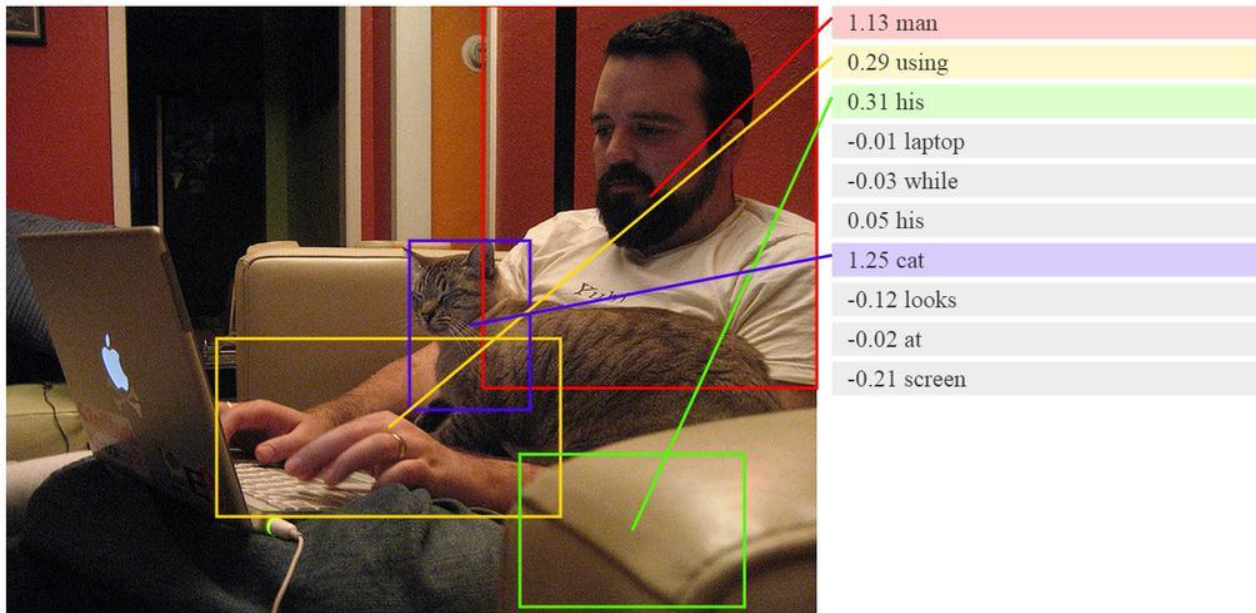

ТЕКСТОВОЕ ОПИСАНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

ВЫПОЛНИЛИ:
МАКАРОВА О.
ПРИБЫТКИНА Д.

ЗАДАЧА

Описать содержание картинок связными предложениями. В качестве входных данных используется изображение, а на выходе получаем описательное предложение – распознанные предметы и объекты. В своей работе мы использовали систему Neuraltalk.

NEURALTALK

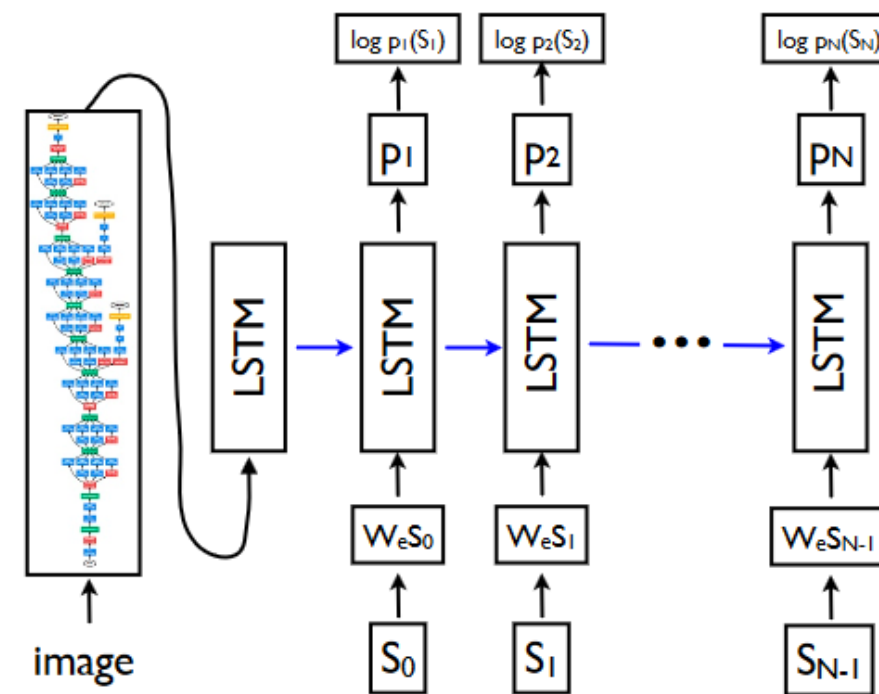


NeuralTalk — это приложение, работающее по принципу нейронных сетей, основанное на разработках Стендфордского Университета и Google.

NeuralTalk способно проанализировать комплексное изображение и точно определить, что на нём происходит, описав всё увиденное разговорным человеческим языком.

АРХИТЕКТУРА

Модель основана на сверточной нейронной сети, которая работает с изображением в рекуррентной нейронной сети и преобразует его в компактное представление. А уже с помощью рекуррентной нейронной сети генерируется соответствующее предложение.



РАБОТА СИСТЕМЫ

Задача разбивается на две стадии:

1. Подпрограмма находит на фотографии объекты и классифицирует их на основе слов из существующего словаря.
2. Другая подпрограмма подбирает описательные слова к разным регионам фотографии и затем объединяет их в предложения.

Для реализации обеих стадий алгоритма используются искусственные нейронные сети. Они обучаются на подборках фотографий Flickr8K и MS-COCO, в которых почти 150000 снимков были подписаны при помощи краудсорсинговой платформы Amazon Mechanical Turk.

ИСПОЛЬЗОВАННЫЕ НЕЙРОННЫЕ СЕТИ

В обоих случаях для характеристики CNN используется VGGNet. Для генерации предложений был использован метод BeamSearch (последовательное рассмотрение наборов из beam лучших предложений).

1. beam = 20

2. beam = 1 (жадный поиск)

LSTM обучается на данных Cocola с 512 скрытым слоями

ЭКСПЕРИМЕНТЫ

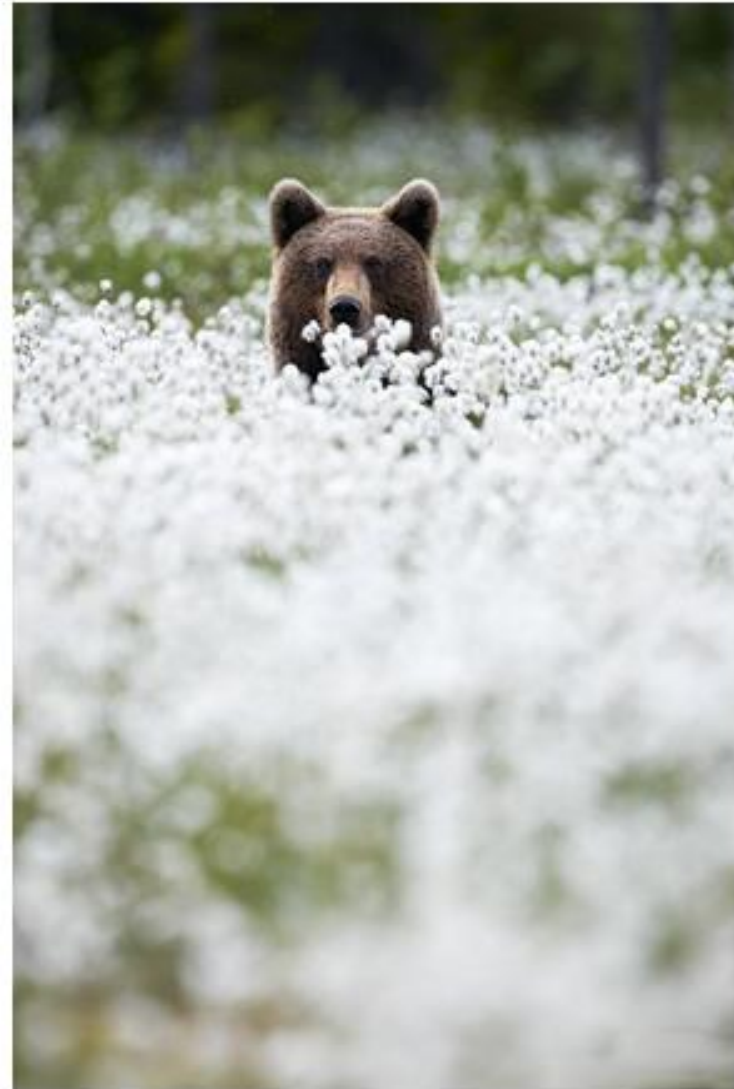
Для оценки используется наборы данных, которые состоят из изображений и предложений на английском языке, описывающие эти изображения. Наборы данных включают в себя 6000 (Flickr8k) и 82783 (MS-COCO) тренировочных файлов.

LOG PROBABILITIES

Дополнительный показатель, вводимый для проверки качества перевода с точки зрения системы. Если этот уровень низкий (например, -10), это означает, что модель путается с изображением и, скорее всего, предоставляет не очень хороший прогноз. И наоборот, более высокие цифры (такие как -7) указывают на то, что модель относительно более уверена в исходе.



(-7.035189) a brown dog is running through a field



(-9.023607) a brown bear is standing in the grass



(-10.103282) a dog is laying on a bed



(-10.288552) a small dog is sitting on a toilet



(-11.486791) a child in a red jacket is standing in the snow



(-10.781403) a man and a child are sitting on a bench



(-11.260886) a white dog is jumping over a red and white fence



(-11.495168) a dog is standing in the grass with a frisbee



(-19.250958) a man in a black shirt and a white hat is sitting on a bench



(-11.956867) a man in a suit and tie is standing in front of a mirror

BLEU

Для оценки работы системы используется показатель, применяемый в машинном переводе для оценки качества сгенерированных предложений – BLEU (Bilingual Evaluation Understudy). Он определяет процент n-грамм (n может быть от 1 до 7), совпавших в машинном переводе и эталонном переводе предложения. Вручную обычно оценивают по 5-балльной шкале два показателя: передача смысла (Adequacy) и гладкость речи (Fluency).

BLEU SCORES

	Flickr8k	MSCOCO
B-1	0.582093	0,649
B-2	0.378414	0,464
B-3	0.189930	0,321

TEST PERPLEXITY

Flickr8k	MS-COCO
15.687797 (vocab size 2538)	11.555093 (vocab size 8791)