

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу

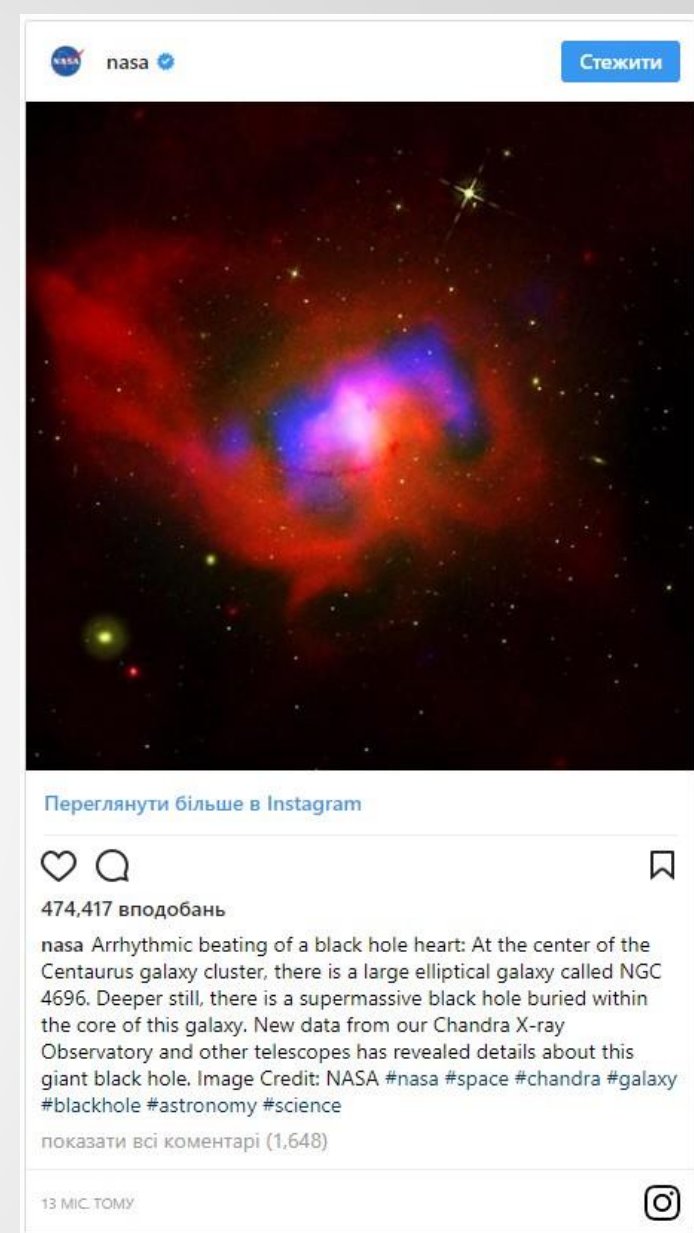
# Система анування зображень на основі байєсівських рекурентних нейронних мереж

Виконав: Гур'янов Олександр, гр. КА-61м

Науковий керівник: к. ф.-м. н., доц. Каніовська І.Ю.

2018 р.

- **Об'єкт дослідження:** системи та методи автоматичного анотування зображень.
- **Предмет дослідження:** байєсівські рекурентні нейронні мережі LSTM, система анотування зображень.
- **Мета роботи:** розробка системи автоматичного анотування зображень, що здатна виділяти об'єкти у зображенні та зв'язки між ними й відображати їх у змістовному підписі природною мовою.
- **Метод дослідження:** методи та моделі комп'ютерного зору та обробки природної мови.
- Задача автоматичного анотування зображень полягає в генеруванні опису довільного зображення природною мовою у вигляді назви або одного чи кількох речень, що відображають його зміст.
- Для розв'язання задачі у роботі пропонується модель, що включає в себе байєсівську рекурентну нейронну мережу.
- На основі навченої моделі було створено програмний продукт – систему автоматичного анотування зображень, що демонструє концепцію та може мати практичні застосування.



# АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

## Автоматичне аотування зображень

- ▶ нещодавно почало активно вивчатись після прогресу в супутніх дослідженнях;
- ▶ може застосовуватись для:
  - індексації зображень;
  - пошуку за зображеннями;
  - пропонування підписів;
  - допомоги людям з вадами зору.
- ▶ є складовою інших задач:
  - створення зв'язних оповідей за набором зображень;
  - підсумовування відео.
- ▶ важливе з теоретичної точки зору як:
  - міждисциплінарна задача – поєднання методів комп'ютерного зору та обробки природної мови;
  - значний крок до машинного розуміння середовища – однієї з цілей в дослідженнях ШІ.

**Тому розробка та вдосконалення моделей аотування зображень є важливими!**

# Постановка задачі

- Математична постановка задачі виглядає наступним чином. Необхідно розв'язати задачу оптимізації відносно параметрів моделі  $\theta$ :

$$\theta^* = \mathop{\text{arg max}}_{\theta} \sum_{(I,S)} \log p(S|I; \theta) = \mathop{\text{arg max}}_{\theta} \sum_{(I,S)} \sum_{t=0}^N \log p(S_t|I; \theta; S_0, \dots, S_{t-1})$$

де  $(I, S)$  – елементи навчальної вибірки.

- Для моделювання умовної ймовірності природно використовувати рекурентну нейронну мережу.



man in black shirt is playing guitar.



construction worker in orange safety vest is working on road.



two young girls are playing with lego toy.



boy is doing backflip on wakeboard.

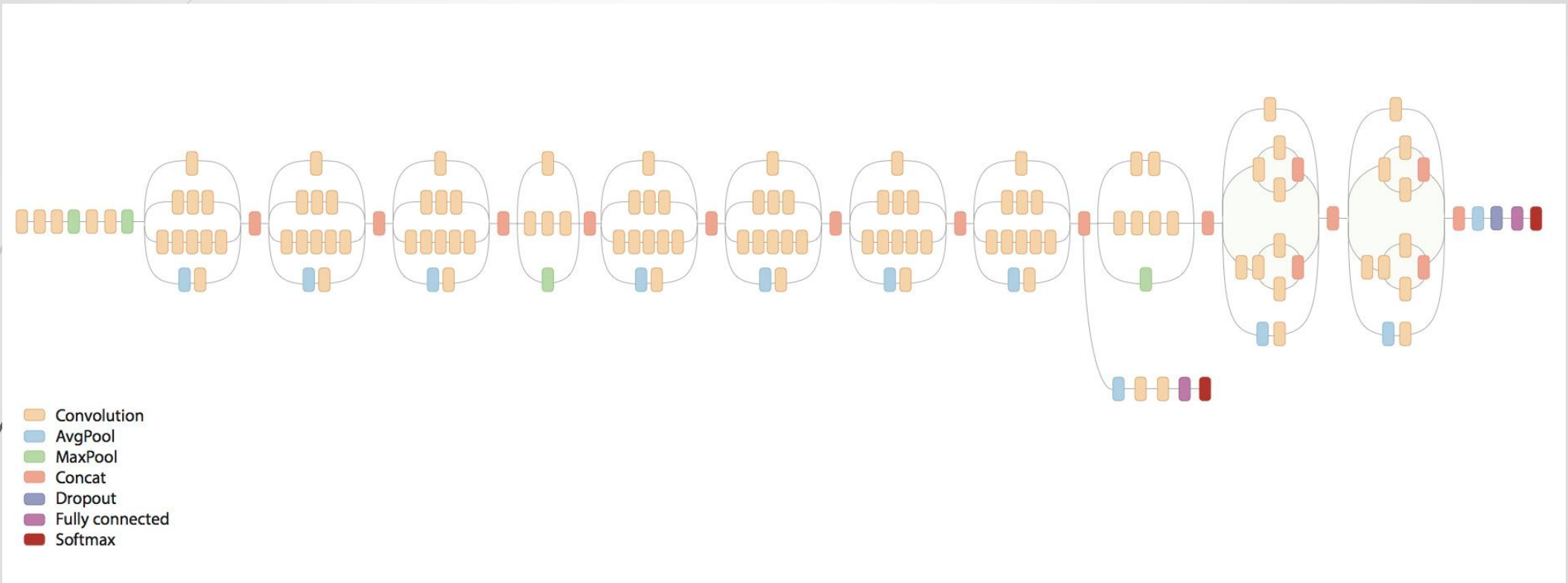
Приклад бажаного результату

# Загальна архітектура моделі

- ▶ На даний час найбільш поширеним підходом до побудови моделі аотування зображень є архітектура кодувальник-декодувальник.
- ▶ Кодувальник – глибинна згорткова нейронна мережа  
вхідне зображення => векторне представлення зображення
- ▶ Декодувальник – рекурентна нейронна мережа  
векторне представлення зображення => аотація

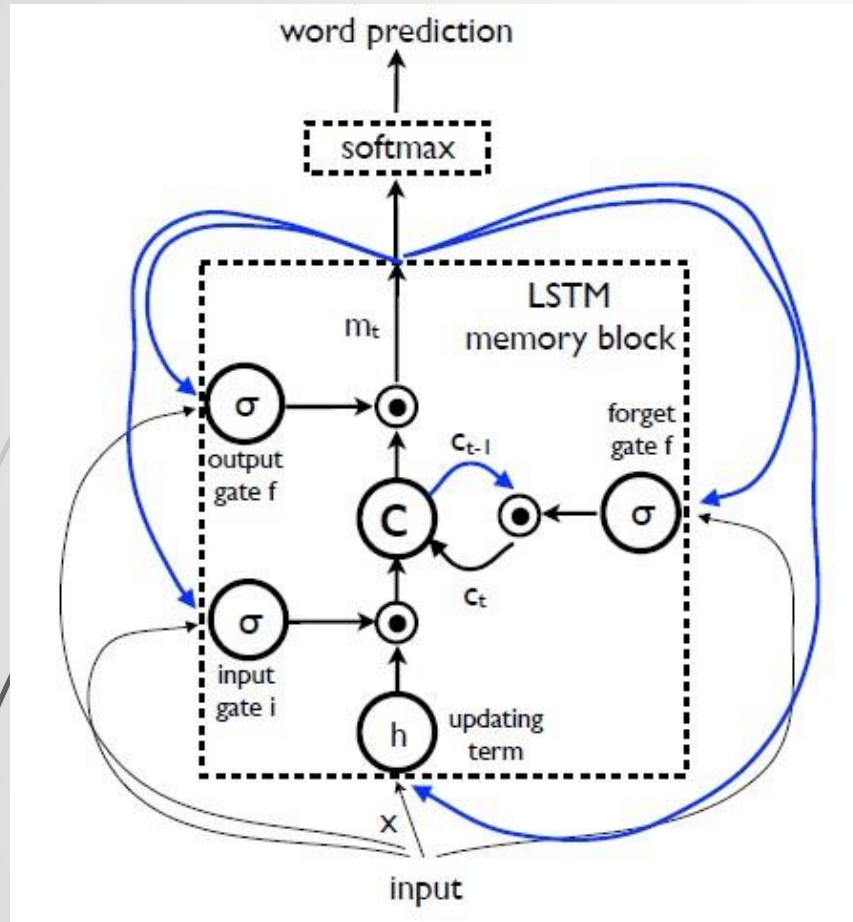
# Глибинна згорткова нейронна мережа

## Inception-v3



- Оптимальний баланс між якістю класифікації, кількістю тренуваних параметрів та швидкістю.
- Розмірність векторного представлення зображення з передостаннього шару: 2048.

# Довга короткочасна пам'ять (LSTM)

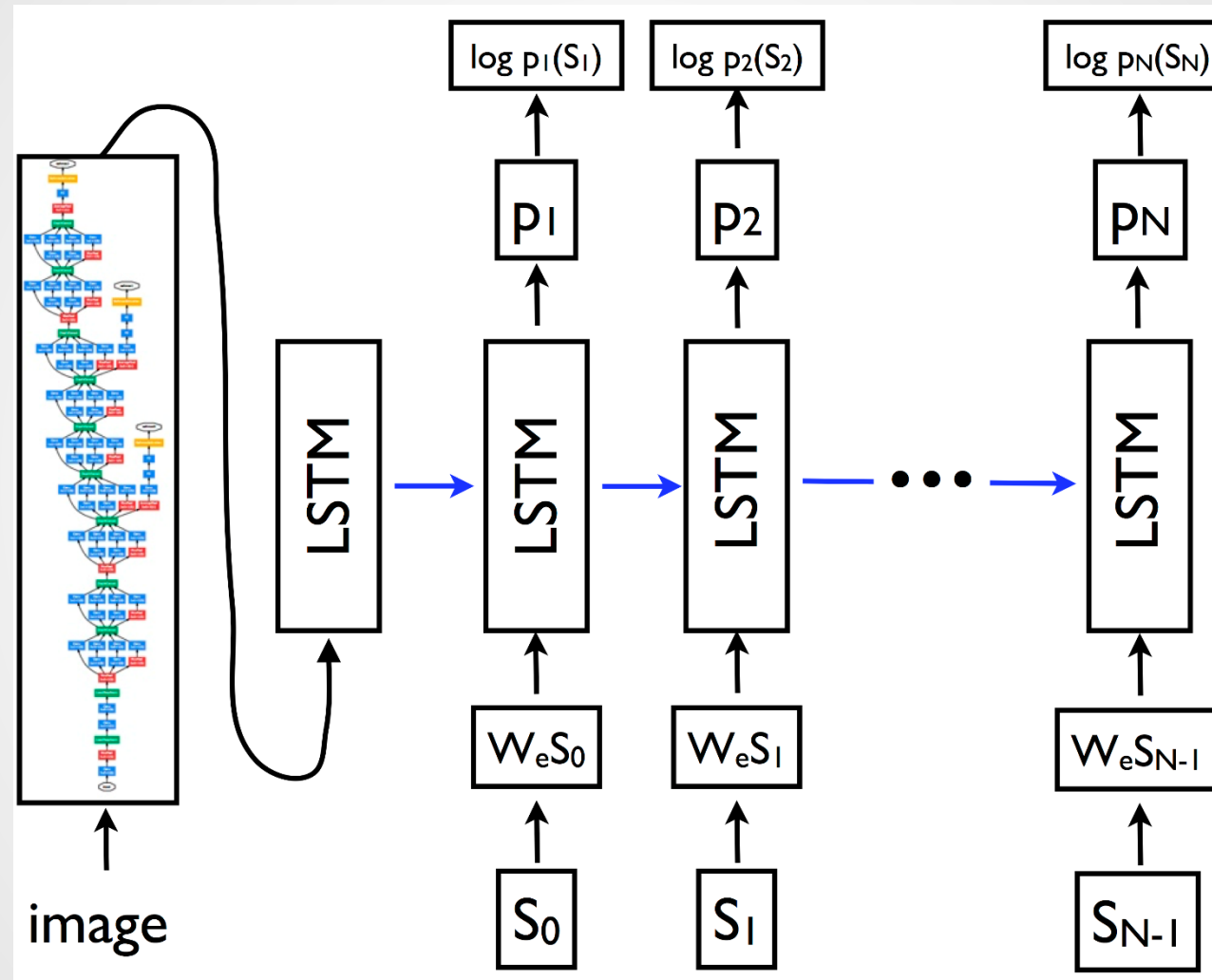


Структура LSTM блоку

- ▶ бореться з проблемою вибухаючих та зникаючих градієнтів при навчанні;
- ▶ має явний вектор пам'яті;
- ▶ успішно застосовується в задачах машинного перекладу та генерування природної мови.

$$\begin{aligned}i_t &= \sigma(W_{ix}x_t + W_{im}m_{t-1}) \\f_t &= \sigma(W_{fx}x_t + W_{fm}m_{t-1}) \\o_t &= \sigma(W_{ox}x_t + W_{om}m_{t-1}) \\c_t &= f_t \circ c_{t-1} + h(W_{cx}x_t + W_{cm}m_{t-1}) \\m_t &= o_t \circ c_t \\p_{t+1} &= \text{Softmax}(m_t)\end{aligned}$$

# Кодувальник-декодувальник



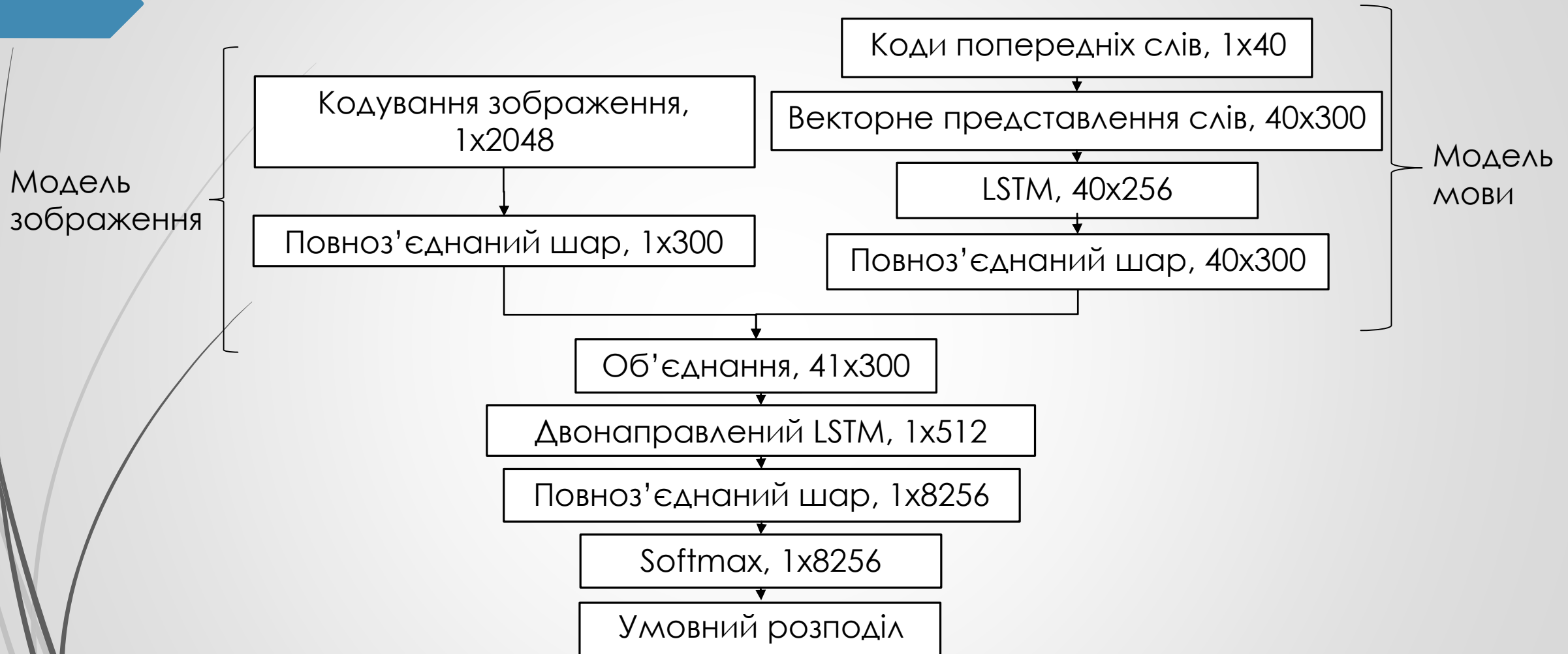
Поєднання моделей комп'ютерного зору та обробки природної мови



# Байєсівські нейронні мережі

- ▶ Ваги мережі є випадковими величинами.
- ▶ Параметрами, що навчаються є параметри спільного апостеріорного розподілу ваг мережі.
- ▶ Для навчання мережі здебільшого адаптується метод зворотного поширення помилки, зокрема як байєсівське зворотне поширення помилки.
- ▶ Переваги:
  - явне врахування параметричної невизначеності;
  - автоматична регуляризація, краща узагальнююча здатність;
  - вихід мережі стає випадковою величиною, можна обчислити її характеристики.
- ▶ Складнощі:
  - вибір апріорного розподілу ваг;   ▪ вибір форми апостеріорного розподілу.

# Структура моделі



Загальна кількість тренуваних ваг: **9115032.**

# Навчальна вибірка

## ► Вимоги до анотацій у навчальній вибірці:

- повно й точно відображають всі об'єкти на зображенні та зв'язки між ними;
- не містять інформації, яку можна дістати лише з контексту;
- не містять власних і специфічних назв.

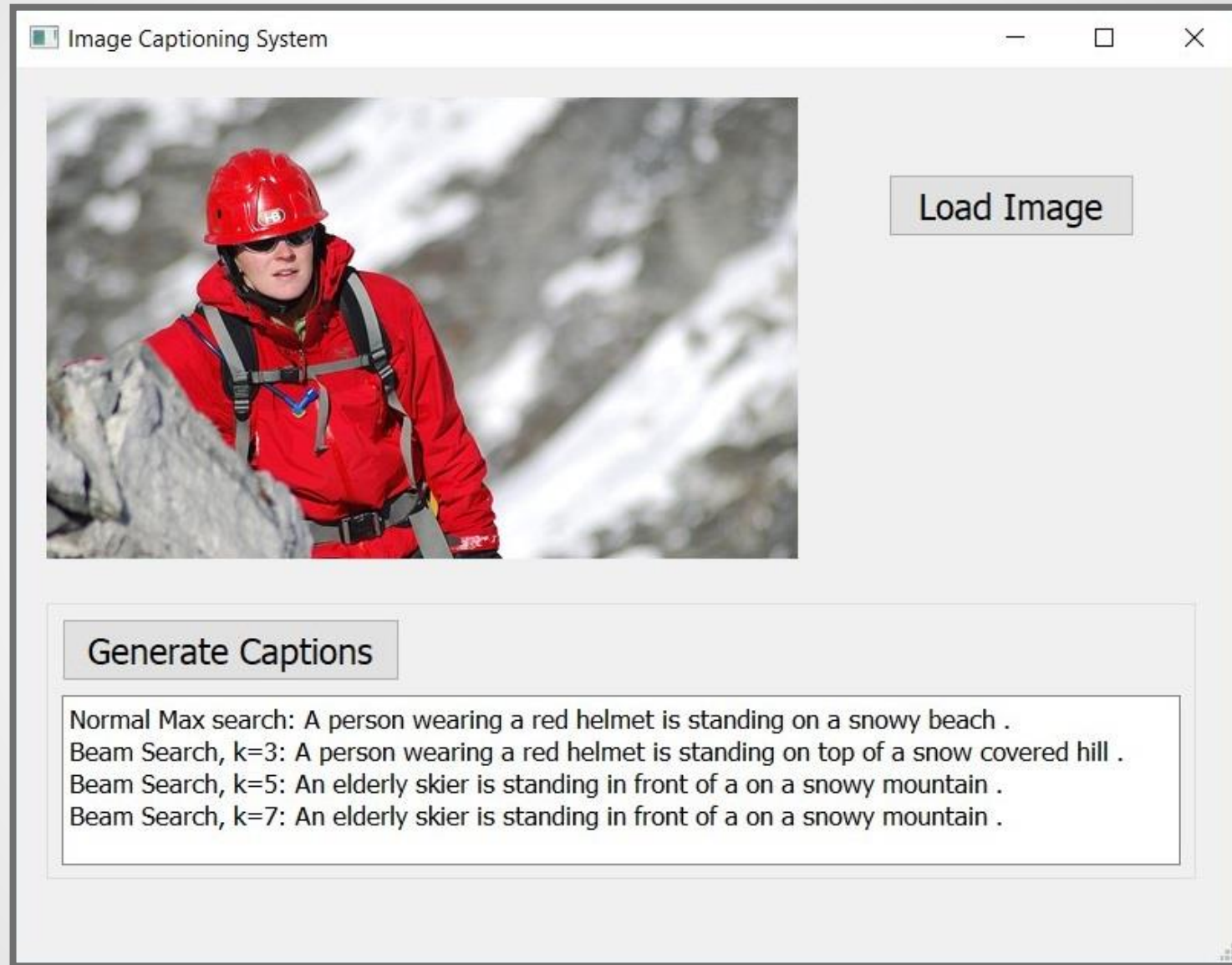
## ► Обрано вибірку Flickr-8k:

- 8000 вручну анотованих зображень;
- по 5 анотацій до кожного зображення;
- середня довжина анотації 11,8 слів;
- 89% анотацій містять дієслова;
- зображення з різних предметних областей.



- 1) A dog in the forest.
- 2) A grey dog explores a fallen tree in the woods.

# Програмний продукт



Інтерфейс системи анотування зображень

# Приклади згенерованих анотацій



Жадібний алгоритм: A hockey player in orange is being tackled.

Променевий пошук,  $k=3$ : An ice hockey player in orange is wearing a red jacket.

Променевий пошук,  $k=5$ : An ice hockey player, wearing orange and black pants.

Променевий пошук,  $k=7$ : An ice hockey player in orange is wearing a red jacket.



Жадібний алгоритм: A young boy is playing basketball.

Променевий пошук,  $k=3$ : A young boy is playing basketball.

Променевий пошук,  $k=5$ : A young boy is playing basketball.

Променевий пошук,  $k=7$ : A young boy is playing basketball.



Жадібний алгоритм: A man sits on a boat at night.

Променевий пошук,  $k=3$ : Two people are sitting in a boat at night.

Променевий пошук,  $k=5$ : A group of people are standing in front of a at night.

Променевий пошук,  $k=7$ : A group of people fishing in a body of water.

# Порівняння зі звичайною моделлю

- ▶ Модель такої самої структури було навчено звичайним методом зворотного поширення помилки.
- ▶ В якості метрики оцінювання якості згенерованих анотацій використовувалась CIDEr.

Модель	Значення метрики CIDEr
Звичайна	78.39
Байєсівська	82.53

- ▶ Байєсівська модель показала кращі результати.


## Висновки

- ▶ запропоновано структуру моделі, дещо спрощену у порівнянні з аналогами;
- ▶ рекурентну LSTM мережу у моделі навчено за допомогою методу байєсівського зворотного поширення помилки, продемонстровано переваги цього підходу;
- ▶ модель показує прийнятні результати, програш новітнім аналогам пояснюється зокрема невеликим обсягом навчальної вибірки зумовленим практичними міркуваннями;
- ▶ на основі моделі побудовано програмний продукт.

# Напрямки подальших досліджень

- ▶ вдосконалення структури моделі;
- ▶ застосування механізмів уваги;
- ▶ дослідження інших форм апостеріорних розподілів при байєсівському навчанні;
- ▶ навчання моделі на більш різноманітних вибірках більшого обсягу.





# Впровадження результатів дипломної роботи

- ▶ За результатами роботи подано статтю «Анотування зображень на основі байєсівських нейронних мереж» до університетського наукового електронного збірника «Системні науки та кібернетика» – 2018. – №6.
- ▶ Побудована математична модель впроваджується в програмному продукті ТОВ «ВІМАС Технології» для вивчення іноземної мови іноземцями на основі технологій розпізнавання мовлення від SRI International.



Дякую за увагу!