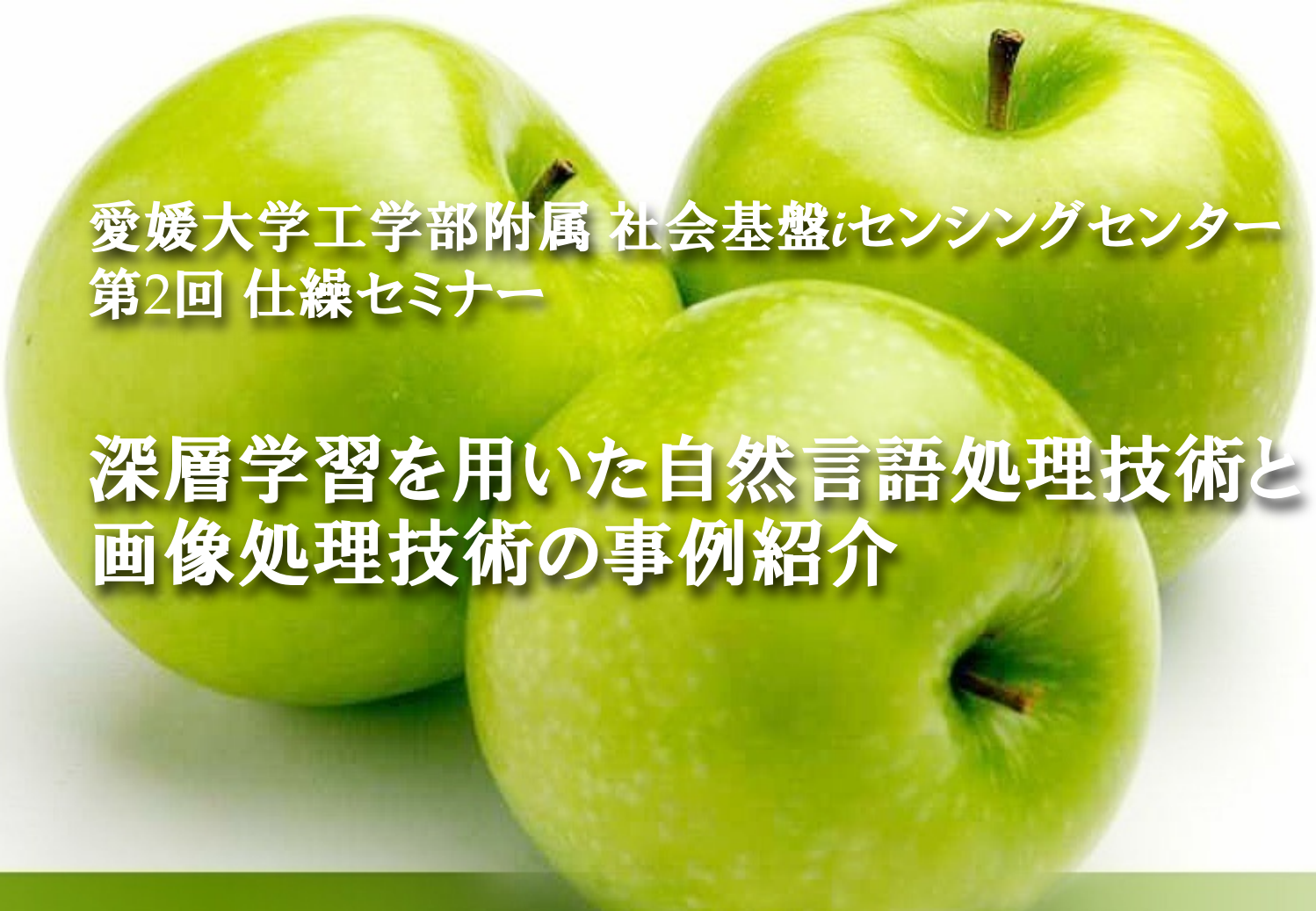


Three green apples are arranged in a cluster, with one in the foreground and two slightly behind it. They are set against a white background.

愛媛大学工学部附属 社会基盤iセンシングセンター
第2回 仕繰セミナー

深層学習を用いた自然言語処理技術と 画像処理技術の事例紹介

愛媛大学 大学院理工学研究科 電子情報工学専攻/
社会基盤iセンシングセンター アナリシス部門
二宮 崇

ninomiya@cs.ehime-u.ac.jp

2020年6月19日

人工知能分野

人工知能(AI)

探索

プランニング

対話

制約ソルバー

確率推論
ベイジアン
ネットワーク

音声認識

命題論理
一階述語論理

強化学習

画像認識
物体認識

知識表現
オントロジー

ロボティクス

自然言語処理

機械学習
(Machine Learning)

深層学習
(Deep Learning)



機械学習：データから法則を見つけ出す手法

深層学習：ニューラルネットワーク(NN)を複数層重ねたモデル

Stuart Russell, Peter Norvig (2010) **Artificial Intelligence: A Modern Approach, 3rd Edition.**



人工知能の分類と自然言語処理

人工知能(AI)

機械学習(Machine Learning)

深層学習

(Deep Learning,  Hot Topics)

自然言語処理

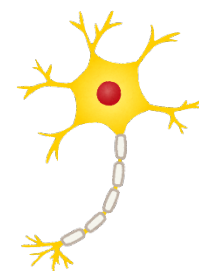


深層学習のインパクト

- **深層学習 = 多層ニューラルネットワーク(NN)による学習**

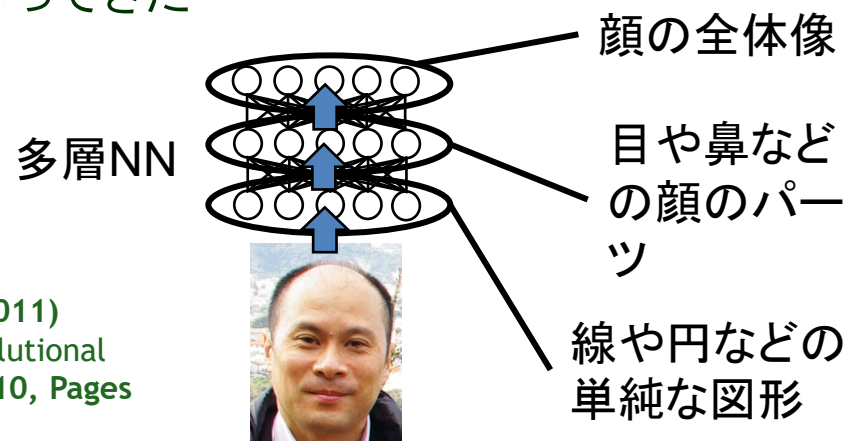
- ニューラルネットワーク

- 人間の脳の神経細胞(ニューロン)の仕組みを模した計算モデル
 - パーセプトロン (Rosenblatt 1958)
 - 畳み込みNN (福島 1980)(LeCun+ 1989)



- **特徴抽出、表現学習**

従来の画像認識では人手で特徴抽出器を作っていたが、深層学習ではこれらの特徴を自動的に学習することがわかってきた



Honglak Lee, Roger Grosse, Rajesh Ranganath, Andrew Y. Ng (2011)
Unsupervised Learning of Hierarchical Representations with Convolutional
Deep Belief Networks, *Communications of the ACM*, Vol. 54 No. 10, Pages
95-103

深層学習のインパクト

- 深層学習の圧倒的精度

- 2012年のILSVRC (ImageNet Large Scale Visual Recognition Challenge)における画像分類の精度
 - 2010年 71.8% (NEC Labs America, Univ. of Illinois at Urbana-Champaign, Rutgers Univ.)
 - 2011年 74.2% (Xerox Research Center Europe, CIII)
 - 2012年 83.6% (Univ. of Toronto) … この年に深層学習が圧勝
 - 2013年 88.3% (Clarifai)
 - 2014年 93.3% (Google)
 - 2015年 96.4% (MSRA) … 人間の分類精度(95%)を超えたと言われる
 - 2016年 97.0% (公安調査庁第3研究所, 中国)
 - 2017年 97.75% (Momenta, Univ. of Oxford)

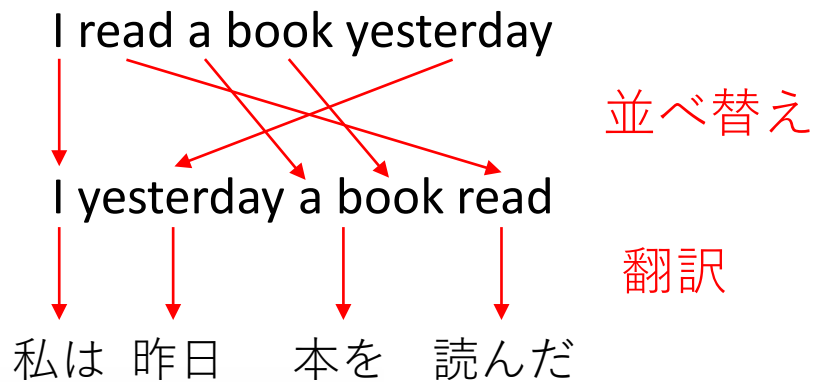
Olga Russakovsky+ (2015) ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, Volume 115, Issue3, pp 211-252
<http://image-net.org/challenges/LSVRC/2017/index.php>



深層学習のインパクト

従来の機械翻訳 「統計的機械翻訳」

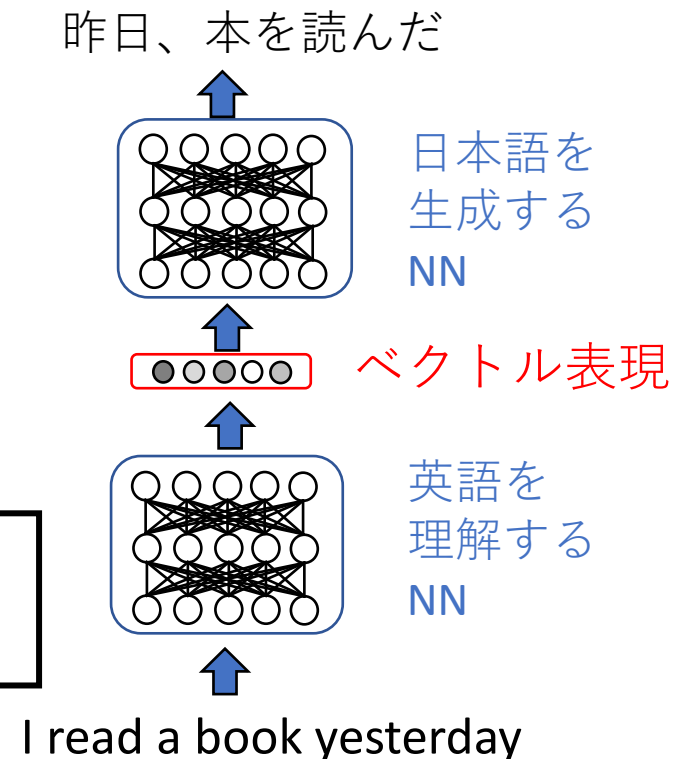
- 並べ替え確率と翻訳確率による確率モデル



従来の統計的機械翻訳の性能を抜いて
人間が行う翻訳にかなり近くなった

深層学習による機械翻訳 「ニューラル機械翻訳」

- ベクトル表現を中間表現とするニューラルネットワーク(NN)



機械学習

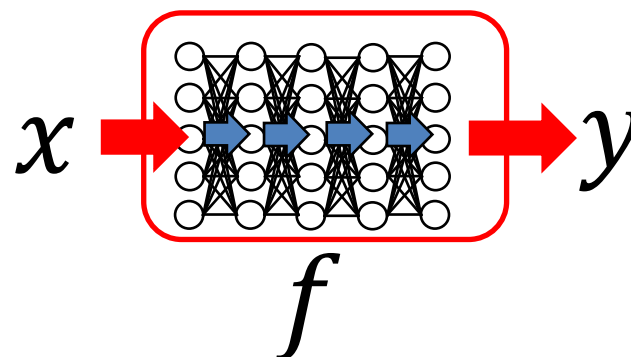
- 機械学習 = データから関数を学習する学問分野

大量のデータ



➡ $f(x)$

- 深層学習も機械学習の一種



機械学習

- データから関数を学習

- 関数は、入力(x)と出力(y)の関係を表す

$$x \xrightarrow{f} y$$

- 大量の入出力ペア(x, y)の集まり(データ)から関数 f を予測

大量のデータ

(x_1, y_1)
 (x_2, y_2)
 (x_3, y_3)
 $\vdots$
 (x_n, y_n)

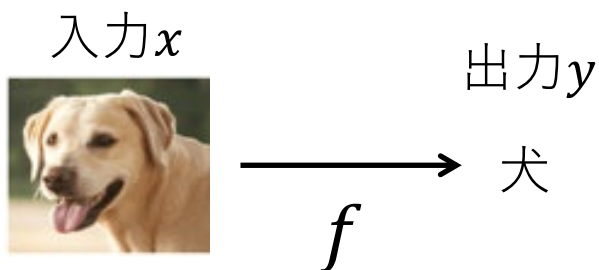
学習

f



機械学習

例：犬猫判別機 $f(x)$



大量のデータから関数 f を学習する $f: \text{画像} \rightarrow \{\text{犬/猫}\}$



機械学習

- データから関数を学習

- 入力 $x = (x_1, x_2, \dots, x_m)$ ← m 次元ベクトル
(画像の場合は、赤、青、緑の画像に対応する3つの行列)
- データ D は入力 x と出力 y のペアの集合
- 関数 f は **重み変数の集合** $w = (w_1, w_2, \dots, w_m)$ からできているとする
例: $f(x) = w_1x_1 + w_2x_2 + \dots + w_mx_m$
 - f は $f_w(x)$ と書くとわかりやすい

- 学習 = データ D に対する誤差を最小にする重み変数 w を求める

データ全体の誤差
$$E(w) = \sum_{(x,y) \in D} (y - f_w(x))^2$$

他にも、マージン最大化、確率最大化による学習がある。
関数の最小値を求める方法は「最適化」という領域で研究されている



教師付き学習と教師無し学習

- **教師付き学習**
 - 入出力ペア(x, y)の集まり(データ)から関数 f を予測
- **教師無し学習**
 - 入力 x の集まり(データ)から関数 f を予測

教師付き学習
(supervised learning)

(x_1, y_1)
 (x_2, y_2)
 (x_3, y_3)
 $\vdots$
 (x_n, y_n)



f

教師無し学習
(unsupervised learning)

x_1
 x_2
 x_3
 $\vdots$
 x_n



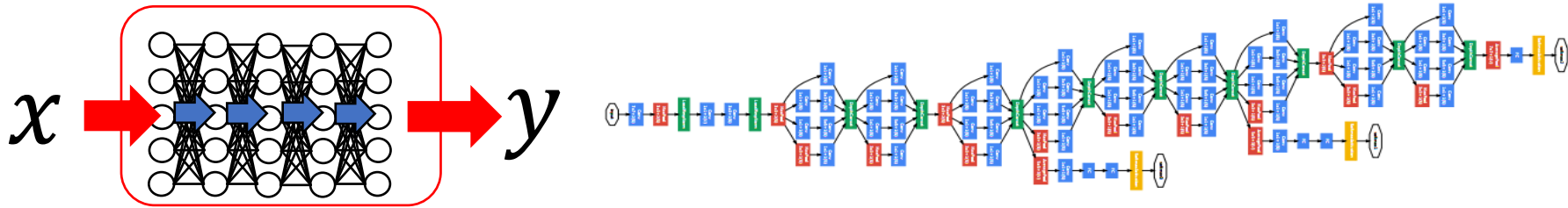
f

正解データ(出力データ)がないので
基本的に難しい

深層学習

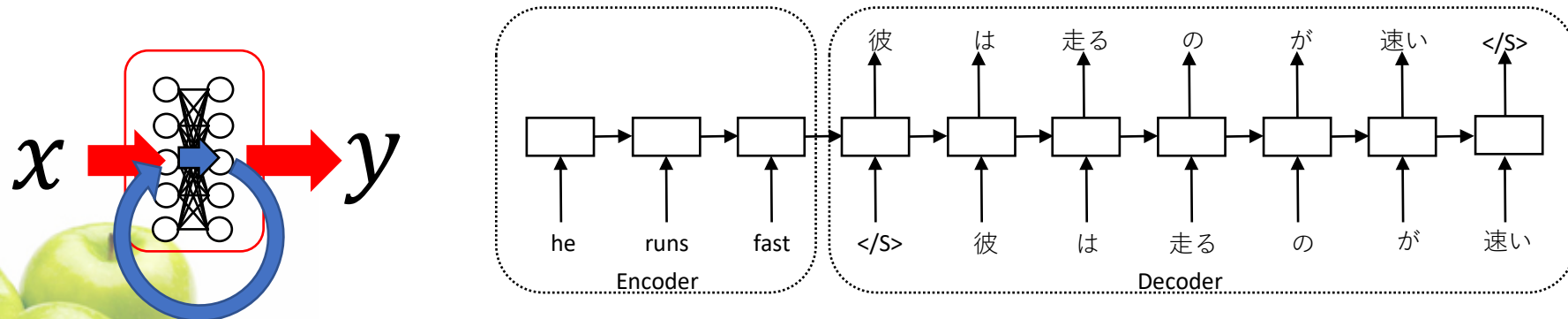
- 関数: 重み付き線形和と非線形変換を多層化したもの

- フィードフォワードニューラルネットワーク



GoogLeNet (22層) ※画像分類に有効なネットワーク

- 再帰型ニューラルネットワーク



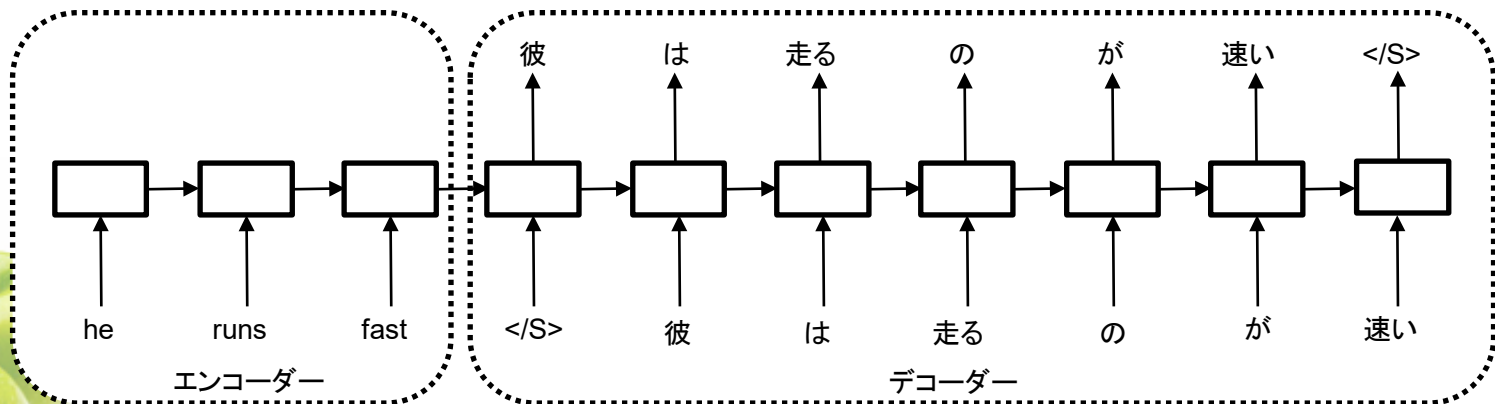
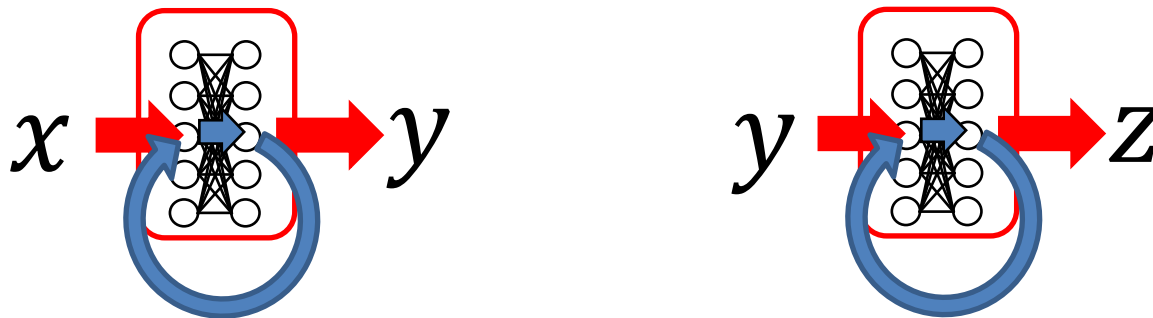
Encoder-Decoderによる機械翻訳

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich (2014) Going Deeper with Convolutions, CVPR 2015

ニューラル機械翻訳

- 2つの再帰型ニューラルネットワークを用いた機械翻訳

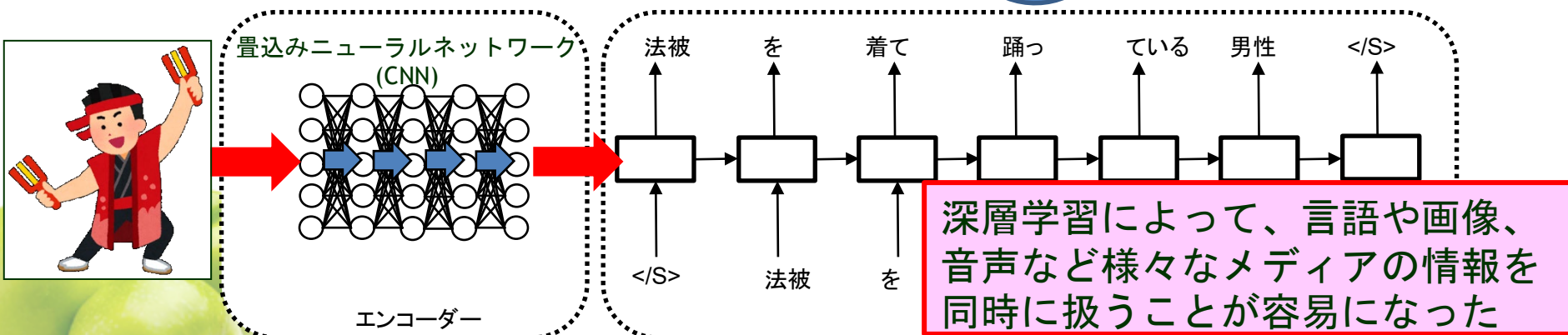
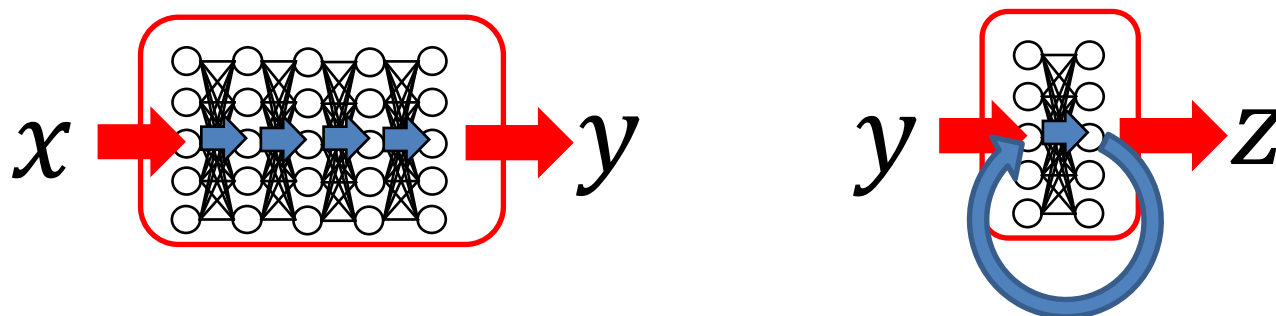
- エンコーダー: 入力文(英語)を中間表現(数百次元のベクトル)に変換
- デコーダー: 中間表現から出力文(日本語)に変換



キャプション生成(説明文生成)

- 畳込みニューラルネットワーク(CNN)と再帰型ニューラルネットワーク(RNN)を用いたキャプション生成

- エンコーダー: 画像を中間表現(数百次元のベクトル)に変換
- デコーダー: 中間表現から出力文(日本語)に変換

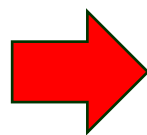


キャプション生成の例(1)

入力画像



生成文

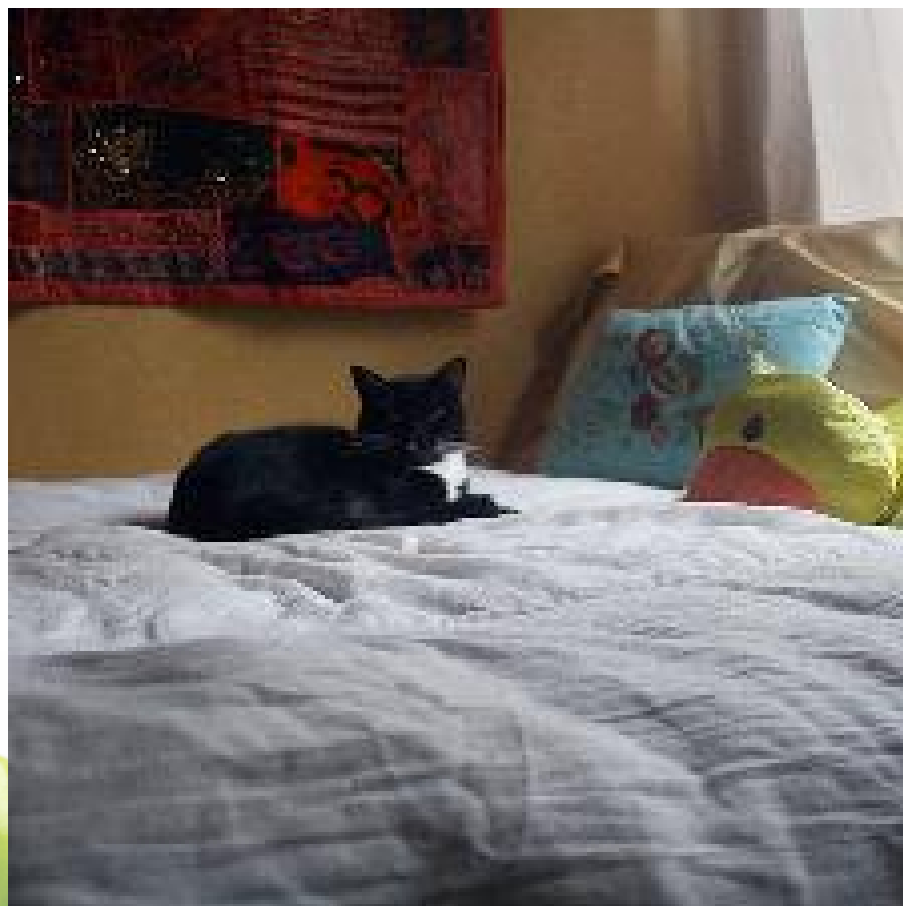


a zebra standing on a snow

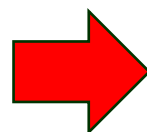


キャプション生成の例(2)

入力画像



生成文



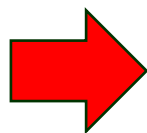
a cat sitting on a bed
with a bed

キャプション生成の例(3)

入力画像



生成文



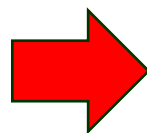
**a bathroom with a sink
with a toilet**

キャプション生成の例(4)

入力画像



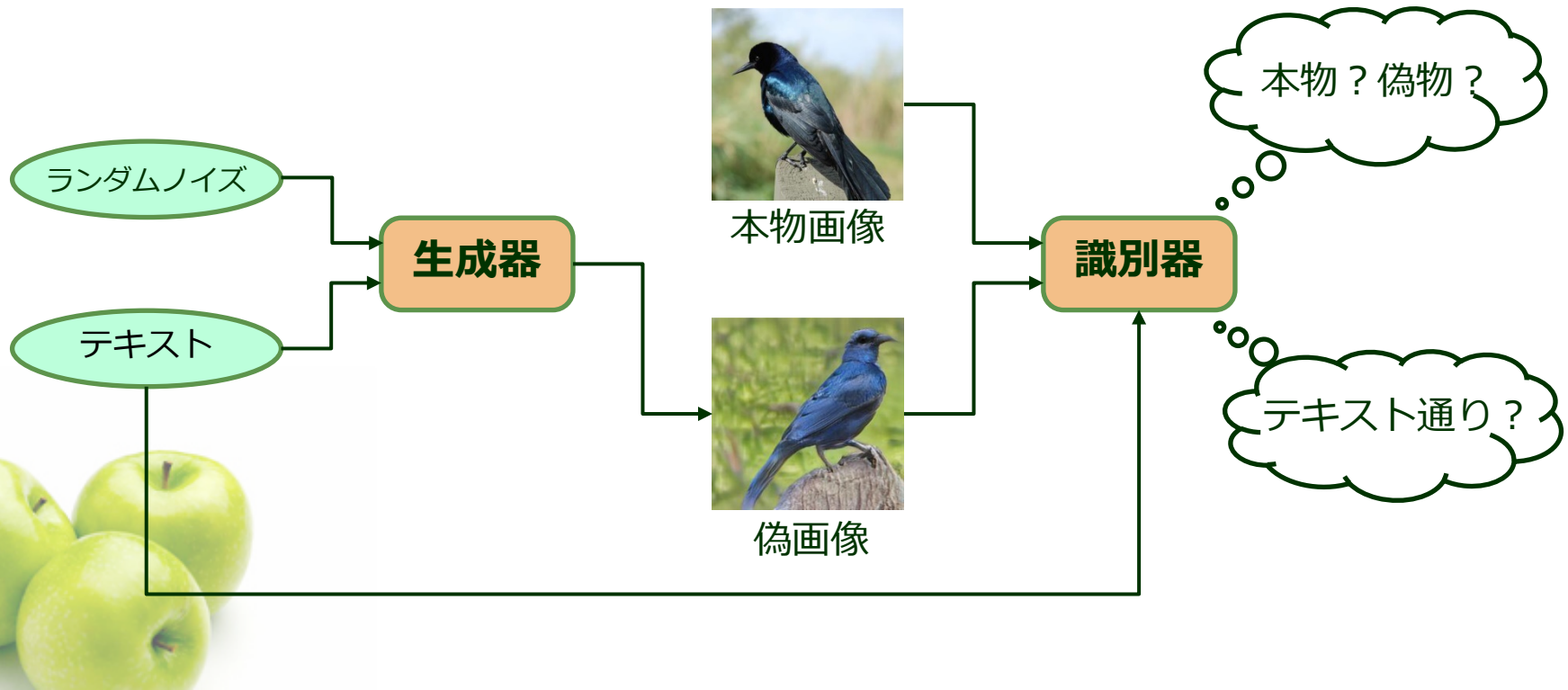
生成文



a man holding a nintendo
wii game

テキスト-画像生成

- 生成器と識別器を互いに競い合わせることで、より本物に近い画像が生成されるように学習(GAN)
 - 生成器：テキストの内容に沿った偽画像を生成
 - 識別器：画像が本物か、テキストの内容に沿っているか判断



テキスト-画像生成の例



英文: Group of people walking
on the heavy snow.

データセットに含まれる本物画像



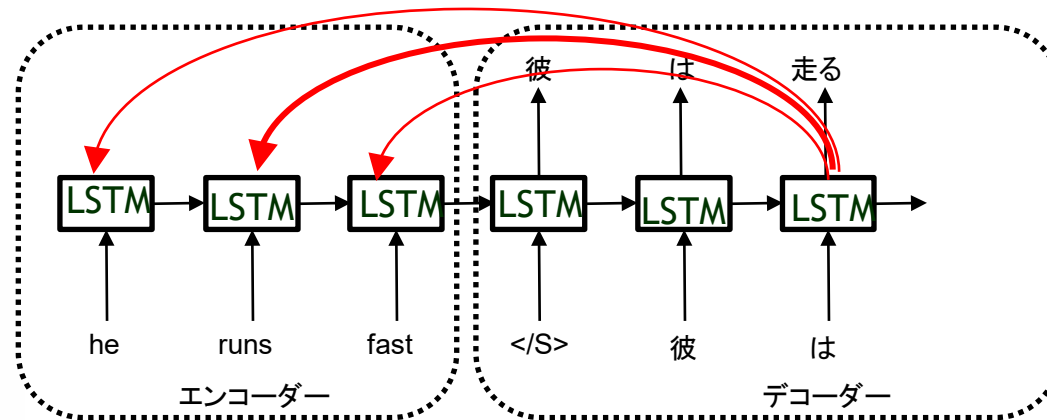
テキスト-画像生成モデルが生成した画像



アテンション付きLSTMエンコーダーデコーダーモデル (Bahdanau+2015)

● アテンション機構とLSTMを用いたニューラル機械翻訳

- LSTM: Long Short-Term Memoryと呼ばれる計算ユニット
- エンコーダー・デコーダー: LSTMを用いた再帰型ニューラルネットワーク
- アテンション機構: エンコーダーの内部状態を参照してデコーダの出力を決定する機構



ニューラル機械翻訳



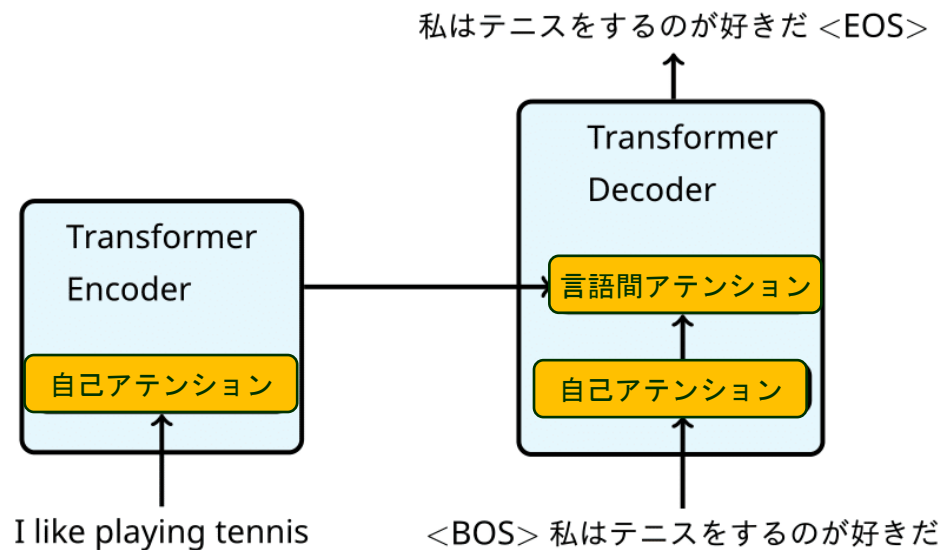
Transformer モデル (Vaswani+2017)

- 自己アテンション機構を用い、文内における単語間の関連の強さを得るのが特徴のニューラル機械翻訳
- 言語間アテンションで原言語・目的言語文間の単語の関連の強さを得ることで、EncoderからDecoderへ入力を行う
- 畳込みニューラルネットワークでも再帰型ニューラルネットワークでもない**新しいタイプのニューラルネットワーク**

➤ 自己アテンション機構の例

(Encoder側)

I like playing tennis

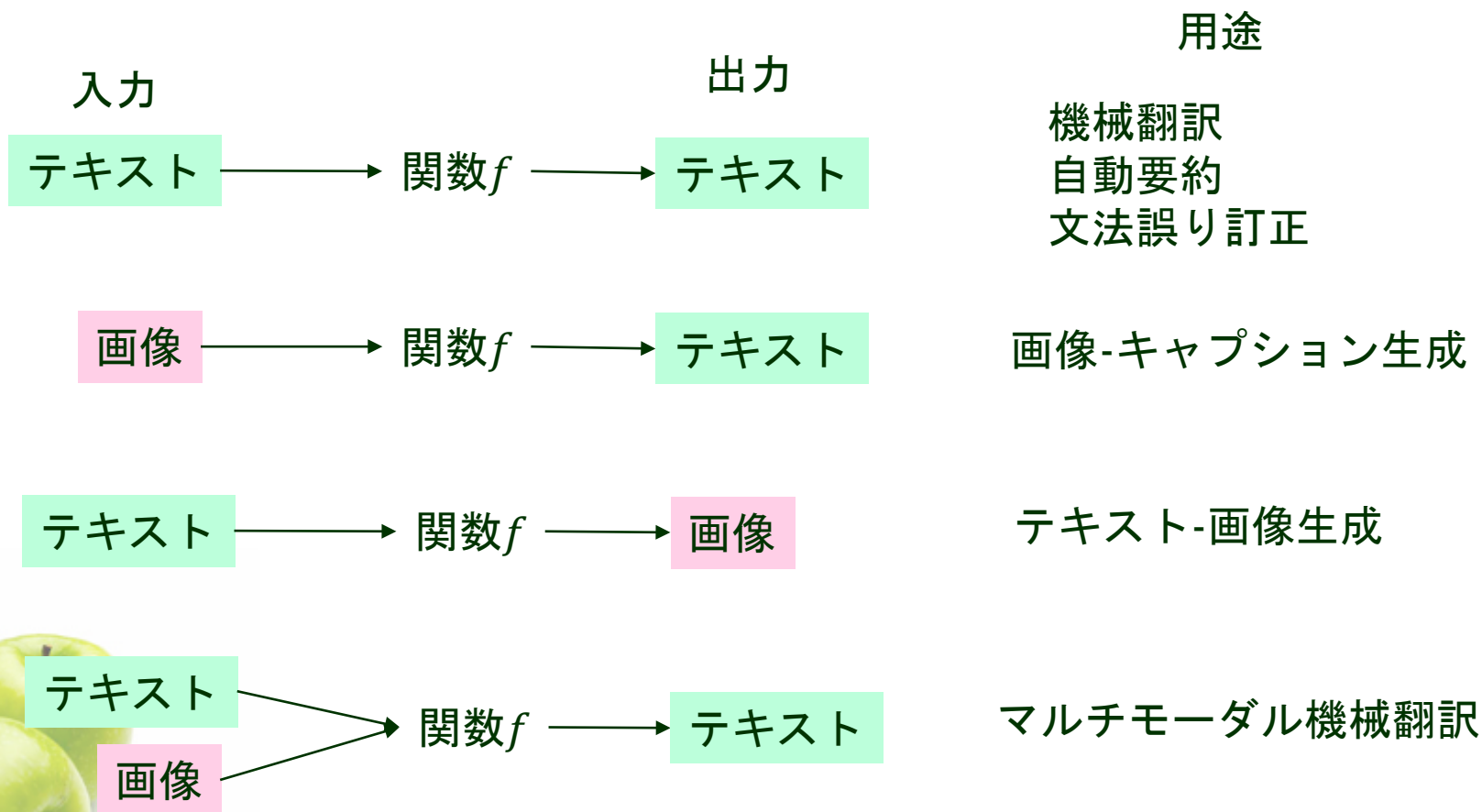


人工知能研究室の研究活動



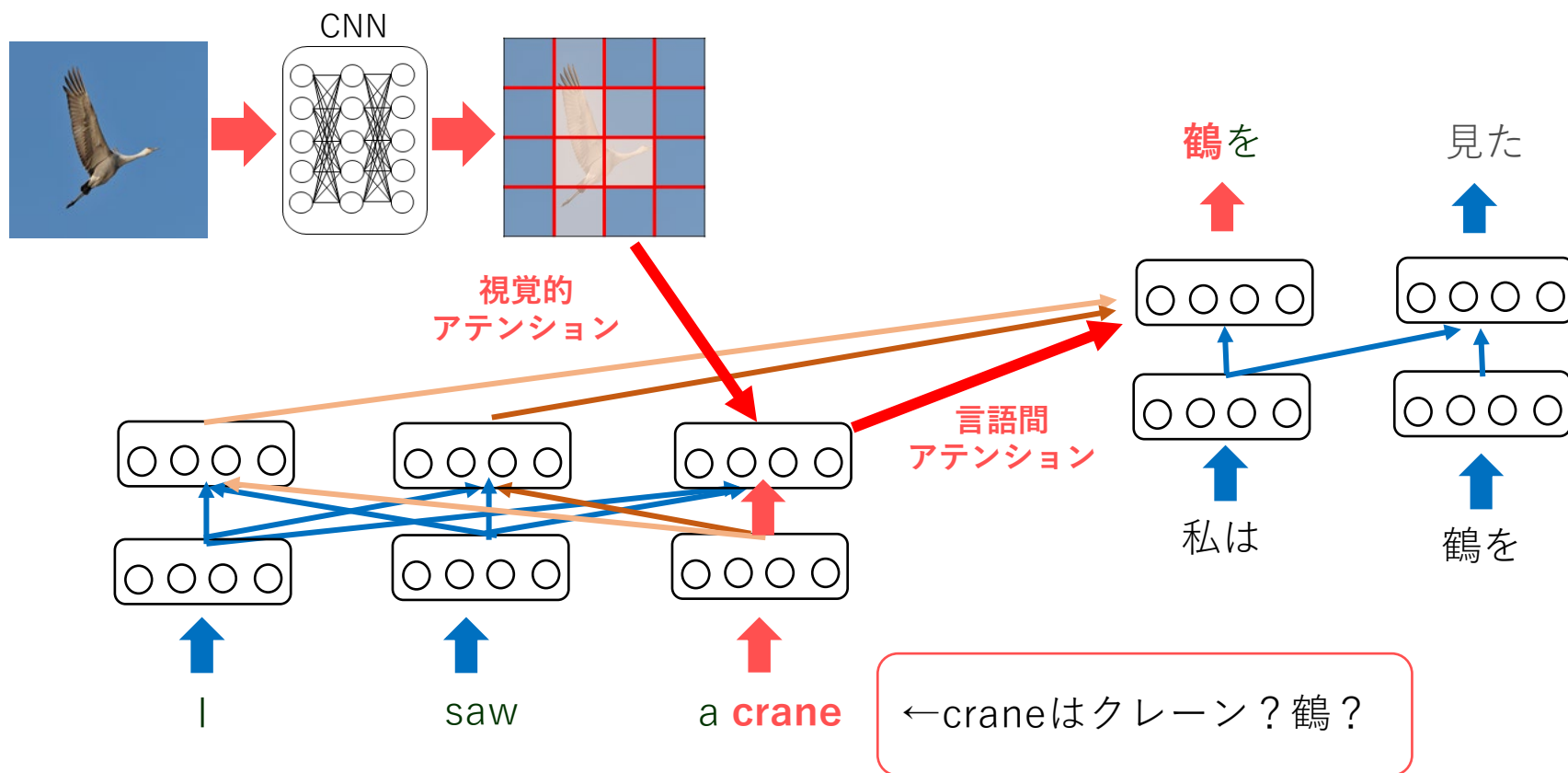
人工知能研究室の研究活動

- 深層学習を使ったマルチモーダル自然言語処理の研究



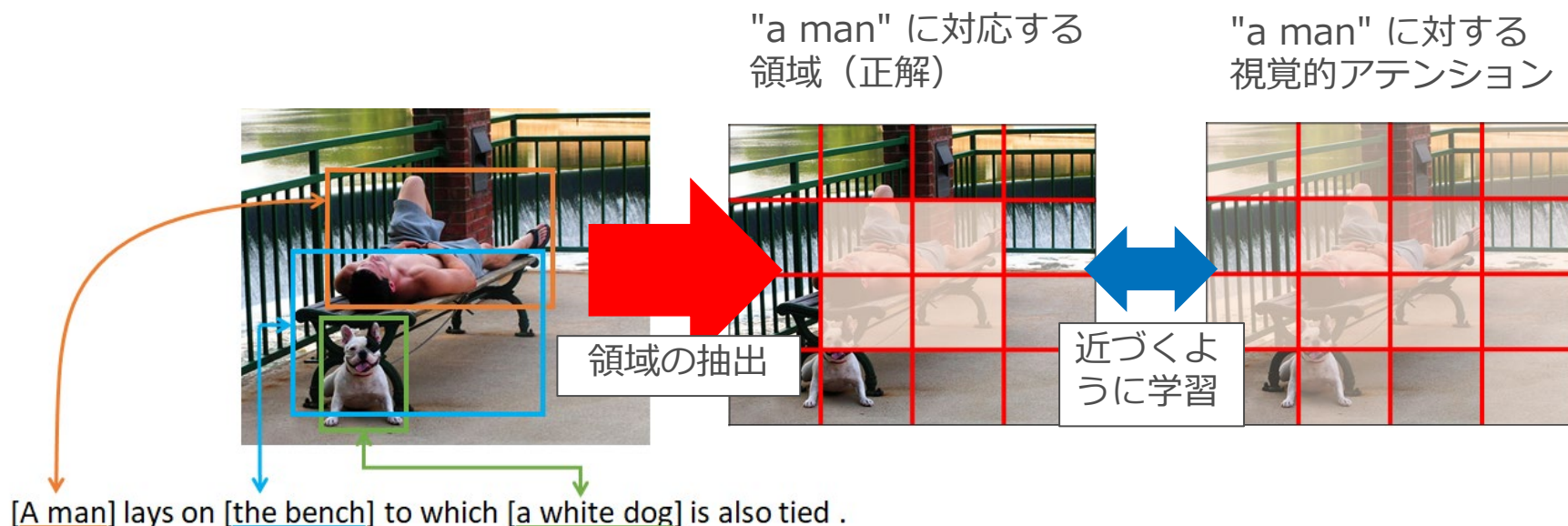
マルチモーダル機械翻訳

- 画像情報を与えることで機械翻訳の精度改善
- 画像内オブジェクトと語/句の対応関係を学習
 - 翻訳タスクを介した画像と語/句のグラウンディング



視覚的アテンション制約

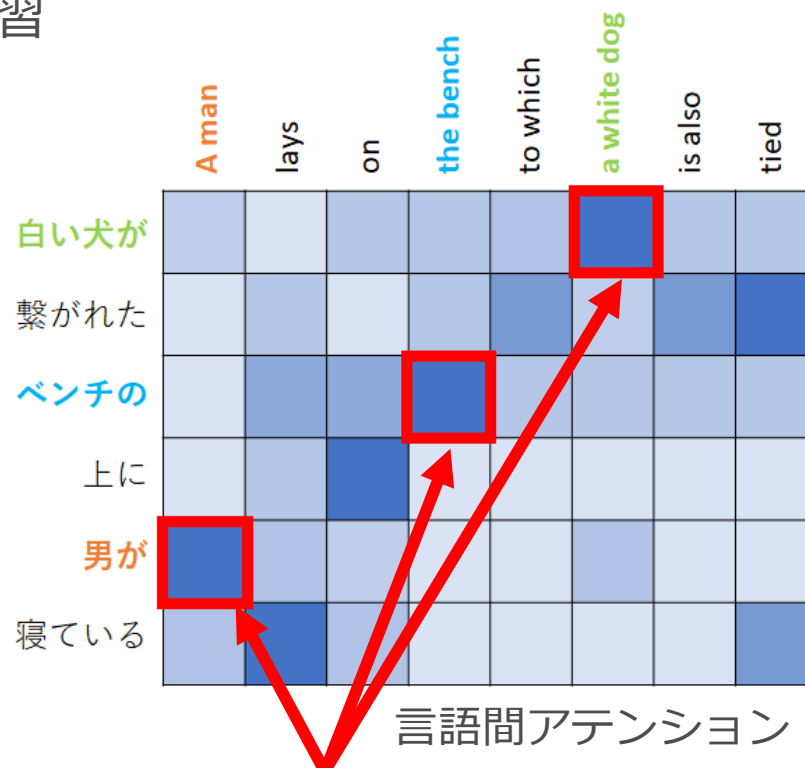
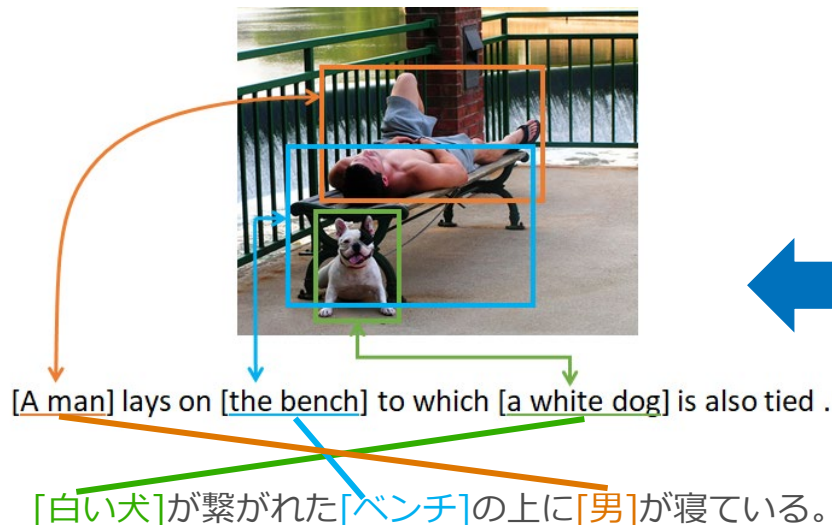
- 視覚的アテンションの教師付き学習
- 画像内の物体と翻訳元文の単語との対応関係を抽出して、視覚的アテンション機構に制約を与えて学習



各単語/句に対する正解領域と視覚的アテンションの領域が近づくように学習する

言語間アテンション制約

- 言語間アテンションの教師付き学習
- 対応関係にある翻訳元単語と翻訳先単語を抽出し、言語間アテンション機構に制約を与えて学習



対応する単語間のアテンションスコアが高くなるように学習する

マルチモーダル機械翻訳の実験

実験設定

- Multi30k(英→日)
 - 訓練データ: 59,516 文
 - 開発データ: 2,017 文
 - テストデータ: 2,000 文
- 画像エンコーダ: ResNet50
- ニューラル機械翻訳: Transformer

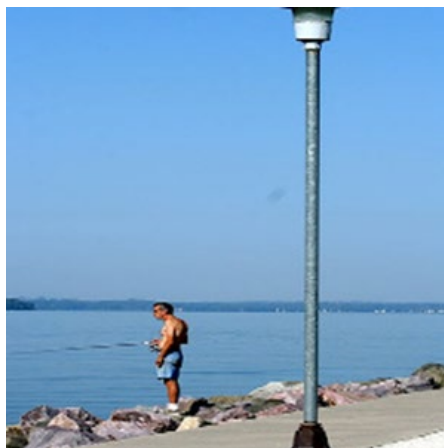
実験結果

モデル		BLEU(%)
従来手法 (教師なしアテンション)		44.11
提案手法 (教師付きアテンション)	視覚的アテンション制約	43.89 (↓0.22)
	言語間アテンション制約	44.61 (↑0.50)
	視覚的アテンション制約 + 言語間アテンション制約	44.98 (↑0.87)

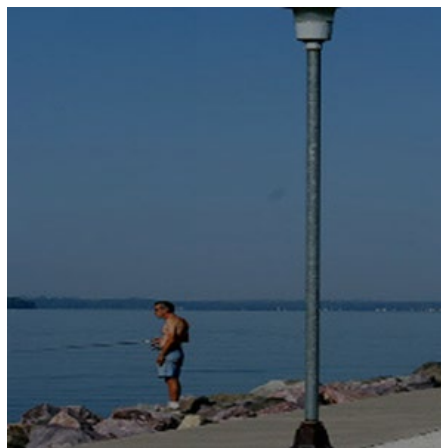
マルチモーダル機械翻訳の例

視覚的アテンションの比較例 (man)

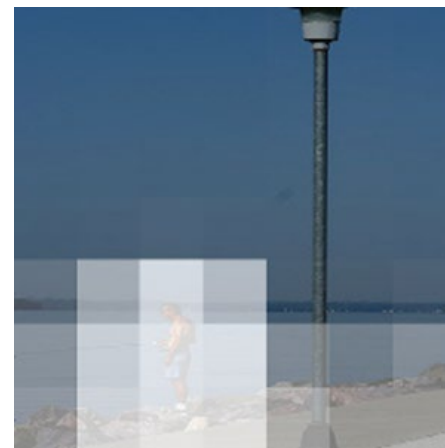
A shirtless **man** is fishing, standing on large rocks that rest next to a boardwalk with a tall lamp post in the forefront.



原画像



従来手法
(教師なしアテンション)



提案手法
(教師付きアテンション)

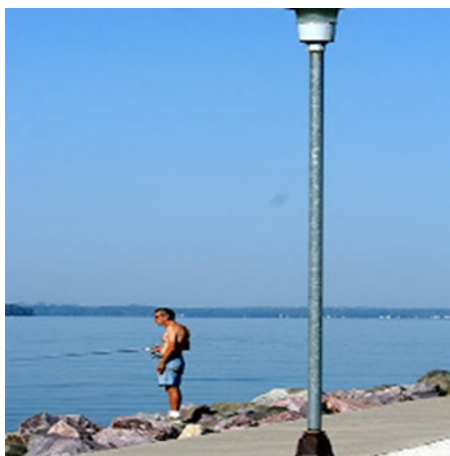
manに対する視覚的アテンション

従来手法ではアテンションがほとんどかかっていないが、提案手法(教師付きアテンション)では**アテンションが強く正確**にかかっていることがわかる

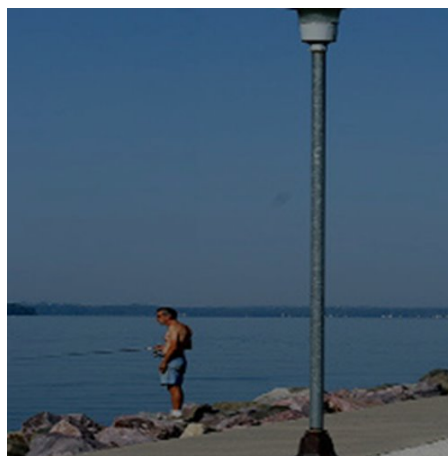
マルチモーダル機械翻訳の例

視覚的アテンションの比較例 (lamp)

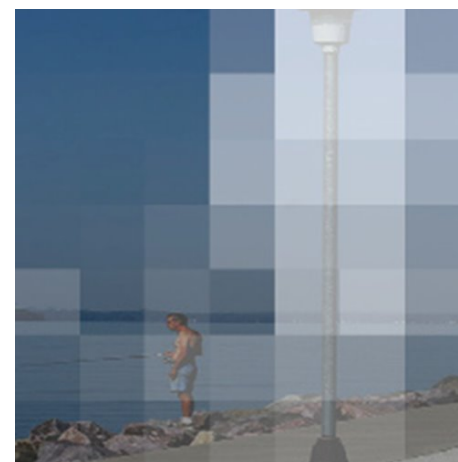
A shirtless man is fishing, standing on large rocks that rest next to a boardwalk with a tall **lamp** post in the forefront.



原画像



従来手法
(教師なしアテンション)



提案手法
(教師付きアテンション)

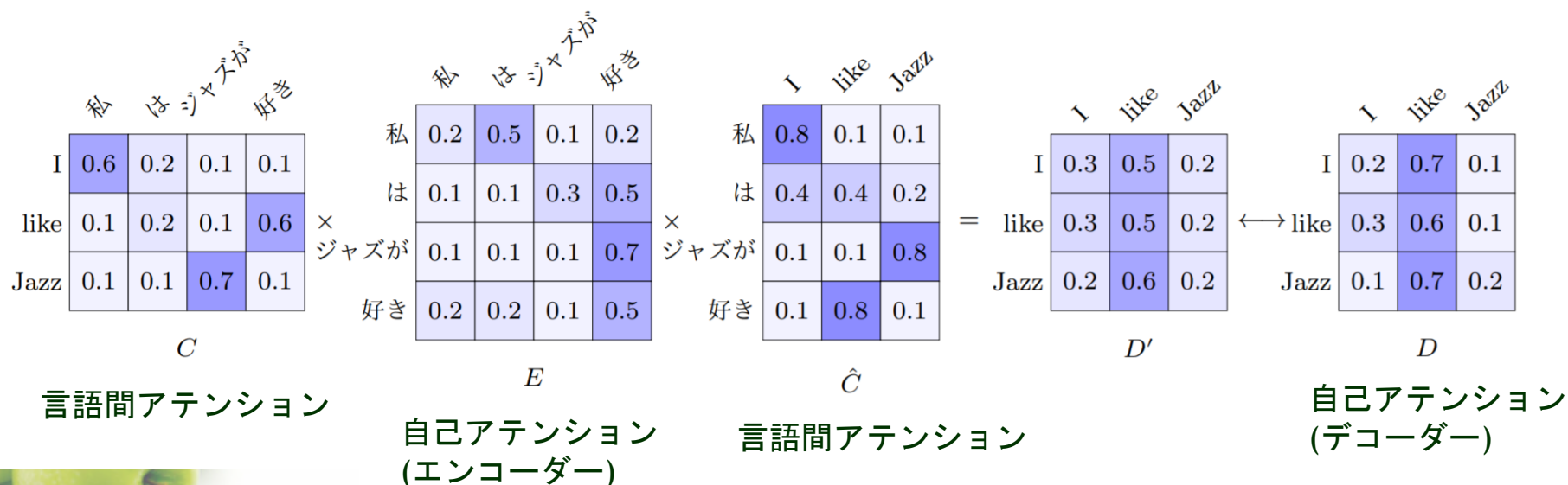
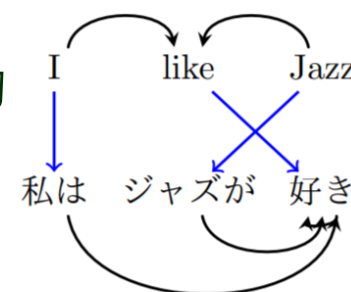
lampに対する視覚的アテンション

manの時と同様、提案手法(教師付きアテンション)では**アテンション**が**lamp**に対し強く**正確**にかかっていることがわかる

同期係り受け制約による機械翻訳

● 自己アテンションと言語間アテンションを用いた同期係り受け制約による機械翻訳の改善

- 自己アテンションを係り受け解析から学習
- 自己アテンションと言語間アテンションの整合性を保つ制約
 - 原言語側で係り受け関係にある単語対は目的言語側でも係り受け関係にある



産学連携研究の事例紹介

- 化合物名に対する構造推定 (理研AIP-富士通との共同研究)

- 化合物名(テキスト)→構造式(テキスト)
- Transformer
- 原子数制約とマルチタスク学習

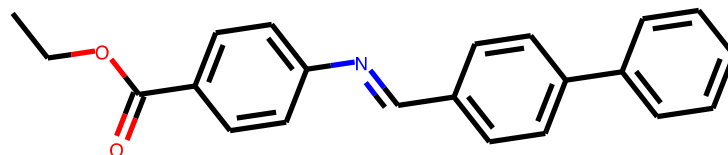
化合物名(テキスト)

Benzoic acid, 4-(((1,1-biphenyl)-4-ylmethylene)amino)-, ethyl ester

関数 f

構造式SMILES(テキスト)

CCOC(=O)c1ccc(N=Cc2ccc(-c3ccccc3)cc2)cc1



産学連携研究の事例紹介

- 化合物名辞書を用いた化合物名解析（理研AIP-富士通との共同研究）

- 文中の化合物名を表す箇所をマークするタスク
- テキスト→タグ列(テキスト)
- BiLSTM-CRF
- 化合物名辞書を用いたマルチタスク学習

テキスト → 関数 f → タグ列(テキスト)

- コールセンター対話ログの自動要約（トランスコスモス社との共同研究）

テキスト(対話ログ) → 関数 f → テキスト(要約)



産学連携研究の事例紹介

- マハラノビス距離を用いた軌道回路の状態基準保全に関する考察（JR西日本との共同研究）

- 鉄道軌道回路における正常状態の統計量を測定
- 正常状態からのマハラノビス距離を計算することで異常状態を検知
- 教師無し学習

軌道回路電圧 → 関数 f

- 機械音を入力とする機械の故障診断

- 教師付き学習
- 教師無し学習

音 → 関数 f → ラベル(故障/正常)



まとめ

- **人工知能研究室の研究紹介**
 - 機械翻訳
 - マルチモーダル機械翻訳
 - キャプション生成
- **産学連携の研究紹介**
 - 化合物名解析、化合物名の構造推定
 - 鉄道軌道回路の異常検出
 - コールセンター対話ログの自動要約
 - 機械音を入力とする機械の故障検出

