

Deep Learning を用いた歯科用器具認識の試み

Image Recognition for Dental Instruments using a Deep Learning

岡 真太郎¹ 野崎一徳² 林 美加子¹

Shintaro Oka¹, Kazunori Nozaki², Mikako Hayashi¹

¹ 大阪大学大学院歯学研究科 口腔分子感染制御学講座 (歯科保存学教室)

¹Department of Restorative Dentistry and Endodontology,
Osaka University Graduate School of Dentistry

² 大阪大学歯学部附属病院 医療情報室

² Division for Medical Information, Osaka University Dental Hospital

Abstract: In general dental clinic, it is required to treat many patients safely in a short time. To do this, it is necessary to shorten the time required for writing the medical records and to prevent incidents by dentists beforehand. Skilled dentists can estimate the treatment contents and incidents from the used dental instruments. Therefore, in order to develop AI for estimating treatment contents by automatic recognition of dental instruments, we conducted experiments and evaluations of object detection using deep learning with bounding boxes.

1. 背景

一般歯科診療においては短時間で多くの患者の処置を安全に行うことが求められるが、診療後のカルテ入力に要する時間が問題の一つとなっている。外科手術時の術野への器具の取り残し防止の観点から、手術室内の器具管理自動化¹⁾が注目されている。一方、一般歯科診療においては器具取り残しのリスクよりはむしろ、診療入力の自動化と異常検知が必要とされている。本研究では、診療内容を推測し、診療入力補助や異常検知を実現するため、その第1段階として、個人情報を含まない歯科用器具を置いたトレーの映像データから、歯科用器具の自動認識を行った。

2. 実験手法

2.1 器具

自動認識の対象となる器具は、大阪大学歯学部附属病院保存科で使用されている歯科用器具のうち、多くの診療内容に共通して用いられる代表的な器具である、デンタルミラー、ピンセット、エクスプローラーの3種類とした。デンタルミラーは口腔内の視診や頬粘膜の圧排に、ピンセットは小さな材料の把持に、エクスプローラーは歯質の触診に用いられる。

2.2 訓練および試験データ

選定した器具をバット (紙トレー) 上に1から3個置き、その様子を直上からカメラ (Apple, iPhone7) で撮影した。これらの画像 (4032×3024、RGB8bit) と、器具を矩形で囲んだ座標情報および正解ラベルを組み合わせたデータを訓練データとした。(図1)

訓練データの画像枚数については、1種類の器具のみがバット上にある画像を各30枚、2種類の器具がバット上で重なっている画像を各20枚、3種類の器具がバット上で重なっている画像60枚とし、計210枚とした。

試験データとしてはバット上で器具が重なっている画像および重なっていない画像を用いた。

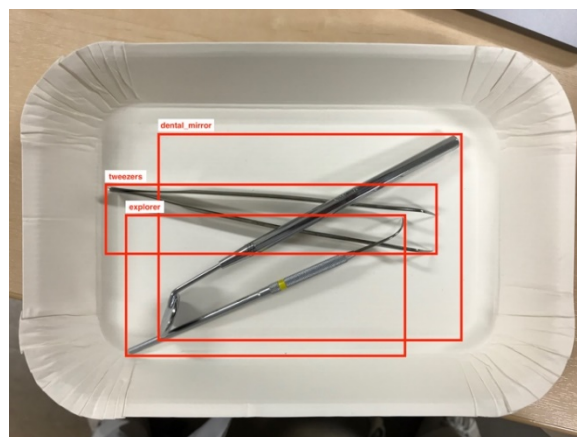


図1：訓練データの例 (3種類の器具)

2.3 Deep Learning

器具認識にあたっては Single Shot MultiBox Detector (SSD)^[2]を用いた。SSD は認識すべき物体が複数ある場合に有効な方法で、識別結果と存在範囲を同時に計算する事ができる。本研究では、学習器、識別器をフリーソフトウェアを用いて実装し、フレームワークとして Tensorflow (ver.1.1.0)、ライブラリとして Keras (ver.2.1.1)、を用いた。

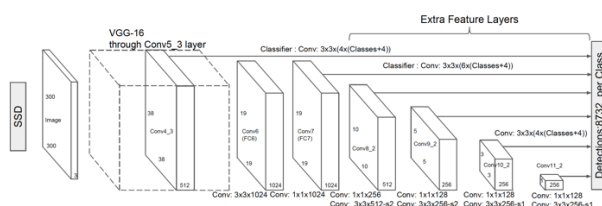


図 2 : VGG16 を改変した学習器 ([2]より引用)

学習器は W. Liu らが提案^[2]する、VGG16 を改変した CNN 構造を用いて ILSVRC CLS-LOC データセット^[3]で事前学習した重みを pool4 まで用い、それ以降の層を今回の訓練データでファインチューニングした。

2.4 評価方法



図 3 : 試験データの例

訓練データと同じ方法でバット上に器具を配置し、図 3 のように角度を変えて撮影した。この組み合わせを訓練データの半数用意し、これらに対して今回ファインチューニングした重みを用いた SSD で物体検出を行った。我々が開発しようとしているシステムの基礎となる器具認識は、器具の位置情報よりも器具の有無が重要となる。そのため器具の存在を示すバウンディングボックスの位置がずれていても、3 枚の検出結果中の 1 枚にでも器具が検出されれば Positive、有無があていれば True と判定し、結果を集計した。

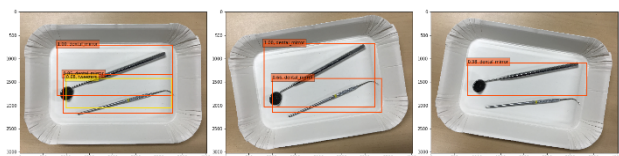


図 4 : SSD による物体検出例

例えば、図 4 左・中・右においてデンタルミラーが検出され、左においてピンセットが誤検出され、全てにおいてエクスプローラーが検出されていない。そのためデンタルミラーの True Positive (TP)が 1、ピ

ンセットの False Positive (FP)が 1、エクスプローラーの False Negative (FN)が 1 となる。

このようにしてバット上に器具が 1 種類ある画像 135(45×3)枚、2 種類ある画像 90(30×3)枚、3 種類ある画像 90(30×3)枚の 3 群について評価を行った。評価項目として、

$$accuracy := \frac{TP+TN}{TP+FP+FN+TN}$$

$$precision := \frac{TP}{TP+FP}, recall := \frac{TP}{TP+FN}, F - measure :=$$

$$\frac{2 \times Recall \times Precision}{Recall + Precision}$$

3. 結果

器具が 1 種類のみバット上に置かれている場合、各器具の検出結果は表 1 のようになった。器具が 2 種類バット上に置かれている場合、各器具の検出結果は表 2 のようになった。器具が 3 種類バット上に置かれている場合、各器具の検出結果は表 3 のようになった。

表 1 : バット上の 1 つの器具の検出結果

	Accuracy	Precision	Recall	F-measure
デンタルミラー	97.8%	93.8%	100%	96.8%
ピンセット	97.8%	93.8%	100%	96.8%
エクスプローラー	97.8%	93.8%	100%	96.8%

表 2 : バット上の 2 つの器具の検出結果

	Accuracy	Precision	Recall	F-measure
デンタルミラー	90.0%	87.0%	100%	93.0%
ピンセット	76.7%	76.0%	95.0%	84.4%
エクスプローラー	70.0%	100%	55.0%	71.0%

表 3 : バット上の 3 つの器具の検出結果

	Accuracy	Precision	Recall	F-measure
デンタルミラー	96.7%	100%	96.7%	98.3%
ピンセット	76.7%	100%	76.7%	86.8%
エクスプローラー	43.3%	100%	43.3%	60.5%

4. 考察

表1について、全ての器具の *Recall* が 100%となったことは、検出漏れがなかったことを示している。しかし、*Precision* が 100%でなかったことは、器具が存在しないのに存在すると誤認識したことを示している。

表2について、デンタルミラーは *Recall* が 100%に対して *Precision* が低く、エキスポローラーは *Precision* が 100%に対して *Recall* が低い。これは、デンタルミラーが検出されやすいが誤検出も多く、エキスポローラーは検出されにくいと誤検出は少ないことを示している。

表3について、*Precision* が全て 100%となっているのは、すべての器具がバット上に存在しているケースについての評価なので、誤検出が存在し得ないためである。つまり *Recall* の値が今回の検出率を正確に表している。

表4：検出結果の総計

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
デンタルミラー	95.2%	94.1%	98.5%	96.2%
ピンセット	85.7%	89.1%	87.7%	88.4%
エキスポローラー	74.3%	97.5%	60.0%	74.3%

全ての検出結果を足し合わせ、各評価項目を計算すると表4のようになった。*Precision* に関しては全ての器具において高い値を示した。これはすなわち器具の誤検出が少ないことを示している。しかし、エキスポローラーの *Recall* が低い値を示していることは、エキスポローラーが検出されにくいことを示している。その原因の一つに、器具のサイズによる影響が挙げられる。SSD では小さい物体の検出精度が低くなることは開発者である W. Liu からも指摘²⁾しており、それに対して SSD を応用した新たな手法などが提案^{4,5,6)}されている。器具が画面上で水平に近い角度で置かれている場合にはエキスポローラー以外でも検出されていないことが多く、これも同様にバウンディングボックスの面積が小さくなるのが原因と考えられる。そのため今回は角度を変えた複数の画像から器具の認識を行うことでこの問題解決を行った。その結果、1つの角度のみの画像で行った器具認識と、3つの角度の画像で行った器具認識の結果(*Accuracy*)を比較すると図5のようになった。

このことから複数の角度で器具を撮影することは *Accuracy* の改善に繋がることが明らかとなった。今

後さらなる精度の向上を行い、器具認識による処置内容の推定の実現に取り組む。

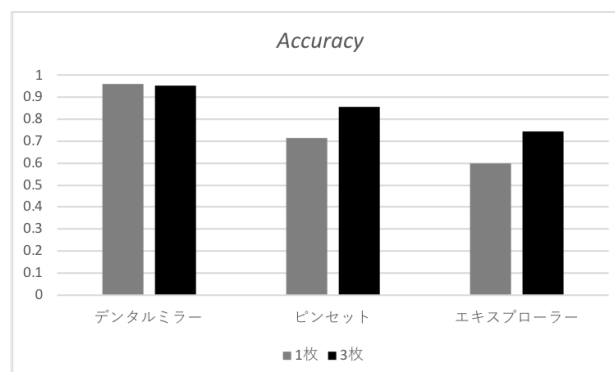


図5：器具認識に用いる画像数の違いによる *Accuracy* の違い

謝辞

本研究の一部は日本電気株式会社共同研究費「スマートデンタルホスピタルに関する研究」の助成による。

参考文献

- [1] Glaser, B., Dänzer, S. & Neumuth, T.: Intra-operative surgical instrument usage detection on a multi-sensor table, *International Journal of Computer Assisted Radiology and Surgery* Volume 10, Issue 3, pp 351–362 (2015)
- [2] Wei Liu, Dragomir Anguelov, Dumitru Erhan.: SSD: Single Shot MultiBox Detector, *European Conference on Computer Vision* (2016)
- [3] Russakovsky, O., Deng, J., Su, H.: Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* (2015)
- [4] Cheng-Yang Fu, Wei Liu, Ananth Ranga.: DSSD : Deconvolutional Single Shot Detector, arXiv, 1701.06659 (2017)
- [5] Guimei Cao, Xuemei Xie, Wenzhe Yang.: Feature-Fused SSD: Fast Detection for Small Objects, arXiv, 1709.05054 (2017)
- [6] Liwen Zheng, Canmiao Fu, Yong Zhao.: Extend the shallow part of Single Shot MultiBox Detector via Convolutional Neural Network, arXiv, 1801.05918 (2018)