

発表スライドに、発表時間の都合で割愛していた参考資料と全データを追加した版
追加部分にはスライド右上に※を付記
また、全データ追加に伴い、スライド38枚目の説明の一部を取り消し線で削除

デンタル画像のマウント位置における 機械学習法の比較と判断根拠の考察

西山 秀昌 1), 織田隆昭 2), 高村真貴 1), 新國 農 1), 勝良剛詞 1), 池
真樹子 1), 坂井幸子 1), 小林太一 1), 曾我麻里恵 1),

Jorge Eduardo Saez Chandia 1), 小椋一朗 3), 林 孝文 1)

1)新潟大学大学院医歯学総合研究科 顎顔面放射線学分野

2)日本歯科大学新潟病院 放射線科

3)日本歯科大学新潟生命歯学部 歯科放射線学講座

はじめに

- 医療へのAIの利用については、海外を含め、厚労省、総務省などからの制約が厳しさを増しており、薬機法を通過させるのは至難の業のように思えるが、**利用者の立場にて考えることは非常に重要**だと思う。
- 多くの学生の視点は、利用者側の視点に近く、**教育という観点から捉えることも非常に重要**だと考える。

使用者責任と提供側説明責任

- 厚労省・使用者責任
 - 医政局医事課長通知(H30.12.19)
- 厚労省・使用者責任と提供側説明責任
 - 薬生機審発0523第2号
- 総務省・説明責任
 - AI活用ガイドライン(R1.8.9)
アカウントビリティ(説明責任)の原則など



画像診断におけるAIと 責任の所在・厚労省・**使用者責任**

- 使用者責任

- 厚生労働省医政局医事課長通知（H30.12.19）

- 「人工知能（AI）を用いた診断・治療支援を行うプログラムを利用して診療を行う場合についても、診断、治療等を行う主体は医師であり、**医師はその最終的な判断の責任を負う**こととなり、当該診療は医師法（昭和23年法律第201号）第17条の医業としておこなわれるものであるので、十分ご留意をいただきたい。」

- 保健医療分野AI開発加速コンソーシアム

- 上記制約を基本に、様々な関連分野で開発を進める方針



画像診断におけるAIと 責任の所在・総務省・説明責任

- AIネットワーク社会推進会議 報告書2019公表
http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html
- AI活用ガイドライン(R1.8.9)
http://www.soumu.go.jp/main_content/000637097.pdf

AI活用原則(10の原則)

- | | |
|--------------|------------------|
| 1) 適正利用の原則 | 6) プライバシーの原則 |
| 2) 適正学習の原則 | 7) 尊厳・自律の原則 |
| 3) 連携の原則 | 8) 公平性の原則 |
| 4) 安全の原則 | 9) 透明性の原則 |
| 5) セキュリティの原則 | 10) アカウンタビリティの原則 |

適正利用の原則



【イ 人間の判断の介在】

- AI サービスプロバイダ及びビジネス利用者は、AI によりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。その場合、人間の判断の介在の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される。
- また、AI によりなされた判断について人間が最終判断をすることが適切とされている場合に、人間が AI と異なる判断をすることが期待できなくなることも想定されることから、説明可能性を有する AI から得られる説明を前提として、人間が判断すべき項目を事前に明確化しておくこと等により、人間の判断の実効性を確保することが期待される²⁰。

統括する家庭内執事ロボットや周辺の AI を含む家電が当該アップデートに適切していないと、（家電同士、または家電とロボットの）相互の判断に齟齬が生じうる（「報告書 2018」別紙 3「AI が想定外の動作を行うなどのおそれ」の事例）。

²⁰ 加えて、人間が確認する AI の判断の適正性を確保するため、他の AI を利用したダブルチェック、AI への入力を摂動させることによる AI 動作の確認などの措置を検討することが望ましい。

透明性の原則



AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入出力等の検証可能性及び判断結果の説明可能性に留意する³⁷。

アカウントビリティの原則



⑩ アカウタバリティの原則

利用者は、ステークホルダに対しアカウタバリティを果たすよう努める。

[ア アカウタバリティを果たす努力]

- AI サービスプロバイダ及びビジネス利用者は、人々と社会から AI への信頼を獲得することができるよう、本ガイドラインの掲げる利活用原則①～⑨の趣旨に鑑み、消費者的利用者、AI の利活用により影響を受ける第三者等に対し、利用する AI の性質及び目的等に照らして、それぞれが有する知識や能力の多寡に応じ、AI システムの特性について情報提供と説明を行うことや、多様なステークホルダとの対話を行うこと等により、相応のアカウタバリティを果たすよう努めることが期待される。

厚生労働省・薬生機審発0523第2号



<https://www.mhlw.go.jp/content/10601000/000515843.pdf>

🔒 mhlw.go.jp/content/

その目的が「診断支援」であるにもかかわらず、医師が**最終診断まで委ねてしまう危険性**を含んでいる。

(3) 責任の所在

人工知能を利用した医療機器においては、その使用に伴う責任の所在が曖昧になることについての懸念が議論されている。特に、支援システム等においては、その目的が「診断支援」であるにもかかわらず、医師が最終診断まで委ねてしまう危険性を含んでいる。よって、支援システムの保守、設計・仕様上の不具合や故障等のトラブル対応においては、通常の医療機器と同様、製造販売業者が責任を負うことは変わらないが、支援システム等の使用目的、使用方法等を使用者に対して明らかにし、使用者に対するトレーニング等を実施する等、適正使用のために必要な方策を検討し、実施の実効性を担保する必要がある。

(6 (4) 項を参照)。なお、平成 30 年 12 月 19 日付け医政医発 1219 第 1 号厚生労働省医政局医事課長通知「人工知能 (AI) を用いた診断、治療等の支援を行うプログラムの利用と医師法第 17 条の規定との関係について」において、当該プログラムを利用して診断、治療を行う主体は医師であり、医師がその最終的な判断の責任を負うことが示されていることも留意されたい。

説明責任＋使用者責任



2) 学習

本評価指標が対象とする支援システム等は、適切な学習データを使用した人工知能の学習によりその目的を達成するために必要となる性能を有することが求められる。よって、対象とする支援システム等の機構や規定された性能等に応じて、以下の項目例を参考にし、必要な項目について内容を明確に示し、またそれらを使用した根拠及び妥当性を示す必要がある。

細部に及ぶ明示が必要

- ・ 学習アルゴリズムとプログラム概要
(教師あり、半教師あり、強化学習、自己学習、その他等を明確に示すこと)
- ・ データ (学習データ、バリデーションデータ^{注1)} 及びテストデータ^{注2)} について、以下を参考に必要な項目を明記すること)

注4) 特に転移学習の場合には、自然画像や別の医用画像、さらには人工画像等を学習データとして用いることが想定される。自然画像を用いた場合には画像中の対象部分、種類及び枚数等、別の医用画像を用いた場合には通常の医用画像を使用した場合と同等の事項、人工画像の場合にはその作成方法等について、その詳細を示すこと。

【重要】機械学習モデルの判断根拠の説明



- 大阪大学 産業科学研究所 原 聡 先生: 機械学習モデルの判断根拠の説明 / 第20回ステアラボ人工知能セミナー (2018/12/21)
 - <https://www.slideshare.net/SatoshiHara3/ss-126157179> (スライド)
 - <https://crash.academy/video/629/1957> (全て閲覧するには、crash.academyに無料会員登録必要)
- ディープラーニングの判断根拠を理解する手法
 - <https://qiita.com/icoxfog417/items/8689f943fd1225e24358>
- 最近の関連論文
 - Sanity Checks for Saliency Maps
Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, Been Kim
 - <https://arxiv.org/pdf/1810.03292.pdf>
- 上記論文の説明関連 / 日本語
 - 深層学習の判断根拠を説明する手法の"不確かさ"を確認した話
 - <https://qiita.com/tsu-gawa/items/e3e4480a7adbc6d1ee17>
 - Sanity Checks for Saliency Maps / データ分析関連のまとめ Hatena Blog (2019-11-24)
 - <https://yhiss.hatenablog.com/entry/2019/11/24/163218>

機械学習のブラックボックスを説明する よりも解釈可能なモデルを使うべき

- Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead
- Cynthia Rudin
- (Submitted on 26 Nov 2018 (v1), last revised 22 Sep 2019 (this version, v3))
- <https://arxiv.org/abs/1811.10154>

機械学習にて鍛えられた 画像診断システムの利用目的は？

• 診断の補助

- 画像診断専門医が利用する。

説明責任

マーケティング
は厳しい？

使用者
責任

診断の補助として
一般歯科医が利用する
のは障壁が高い？

理解できる知識
と、理解しやすい
システムが必要

• おまかせ診断

- 画像診断専門医が利用する。
- 一般歯科医が利用する。

少なくとも学生の教育
は必要？ 現在・大学
院演習では実施中

疑問

- 画像診断支援としての機械学習が盛んに研究・開発されてきているが、**利用者の立場から見て、理解可能なシステム**とは？
- **機械学習の判断根拠が適切かどうか**は、どうすれば分かるのだろうか？
- 透過像での画像診断システムにて、**自然画像（反射光）を用いた「巨大で有名な」システムから転移学習したので安心**といえるのだろうか？
- 機械学習での診断根拠は、**人間が常識的に考えている判断根拠と、同じなのだろうか？**

大量の画像を識別しうる巨大かつ複雑なシステムは
必要なのだろうか？ 転移学習の影響は？

Transfusion: Understanding Transfer Learning for Medical Imaging

医用画像への転移学習を理解する

Maithra Raghu*

Cornell University and Google Brain
maithrar@gmail.com

Chiyuan Zhang*

Google Brain
chiyuan@google.com

Jon Kleinberg[†]

Cornell University
kleinber@cs.cornell.edu

Samy Bengio[†]

Google Brain
bengio@google.com

Dataset	Model Architecture	Random Init	Transfer	Parameters	IMAGENET Top5
RETINA	Resnet-50	96.4% $\pm$ 0.05	96.7% $\pm$ 0.04	23570408	92.% $\pm$ 0.06
RETINA	Inception-v3	96.6% $\pm$ 0.13	96.7% $\pm$ 0.05	22881424	93.9%
RETINA	CBR-LargeT	96.2% $\pm$ 0.04	96.2% $\pm$ 0.04	8532480	77.5% $\pm$ 0.03
RETINA	CBR-LargeW	95.8% $\pm$ 0.04	95.8% $\pm$ 0.05	8432128	75.1% $\pm$ 0.3
RETINA	CBR-Small	95.7% $\pm$ 0.04	95.8% $\pm$ 0.01	2108672	67.6% $\pm$ 0.3
RETINA	CBR-Tiny	95.8% $\pm$ 0.03	95.8% $\pm$ 0.01	1076480	73.5% $\pm$ 0.05

Table 1: **Transfer learning and random initialization perform comparably across both standard IMAGENET architectures and simple, lightweight CNNs for AUCs from diagnosing moderate DR. Both sets of models also have similar AUCs, despite significant differences in size and complexity.** Model performance on DR diagnosis is also not closely correlated with IMAGENET performance, with the small models performing poorly on IMAGENET but very comparably on the medical task.

Model Architecture	Atelectasis	Cardiomegaly	Consolidation	Edema	Pleural Effusion
Resnet-50	79.52 $\pm$ 0.31	75.23 $\pm$ 0.35	85.49 $\pm$ 1.32	88.34 $\pm$ 1.17	88.70 $\pm$ 0.13
Resnet-50 (trans)	79.76 $\pm$ 0.47	74.93 $\pm$ 1.41	84.42 $\pm$ 0.65	88.89 $\pm$ 1.66	88.07 $\pm$ 1.23
CBR-LargeT	81.52 $\pm$ 0.25	74.83 $\pm$ 1.66	88.12 $\pm$ 0.25	87.97 $\pm$ 1.40	88.37 $\pm$ 0.01
CBR-LargeT (trans)	80.89 $\pm$ 1.68	76.84 $\pm$ 0.87	86.15 $\pm$ 0.71	89.03 $\pm$ 0.74	88.44 $\pm$ 0.84
CBR-LargeW	79.79 $\pm$ 0.79	74.63 $\pm$ 0.69	86.71 $\pm$ 1.45	84.80 $\pm$ 0.77	86.53 $\pm$ 0.54
CBR-LargeW (trans)	80.70 $\pm$ 0.31	77.23 $\pm$ 0.84	86.87 $\pm$ 0.33	89.57 $\pm$ 0.34	87.29 $\pm$ 0.69
CBR-Small	80.43 $\pm$ 0.72	74.36 $\pm$ 1.06	88.07 $\pm$ 0.60	86.20 $\pm$ 1.35	86.14 $\pm$ 1.78
CBR-Small (trans)	80.18 $\pm$ 0.85	75.24 $\pm$ 1.43	86.48 $\pm$ 1.13	89.09 $\pm$ 1.04	87.88 $\pm$ 1.01
CBR-Tiny	80.81 $\pm$ 0.55	75.17 $\pm$ 0.73	85.31 $\pm$ 0.82	84.87 $\pm$ 1.13	85.56 $\pm$ 0.89
CBR-Tiny (trans)	80.02 $\pm$ 1.06	75.74 $\pm$ 0.71	84.28 $\pm$ 0.82	89.81 $\pm$ 1.08	87.69 $\pm$ 0.75

Table 2: **Transfer learning provides mixed performance gains on chest x-rays.** Performances (AUC%) of diagnosing different pathologies on the CHEXPART dataset. Again we see that transfer learning does not help significantly, and much smaller models performing comparably.

大量の画像を識別しうる (Resnet50のような)
巨大なシステムは不要。転移学習も不要。
小さな (適切な大きさ) のシステムで十分。

Model	Rand Init	Pretrained
Resnet50	92.2%	94.6%
CBR-LargeT	93.6%	93.9%
CBR-LargeW	93.6%	93.7%

Table 3: **Benefits of transfer learning in the small data regime are largely due to architecture size.** AUCs when training on the RETINA task with only 5000 datapoints. We see a bigger gap between random initialization and transfer learning for Resnet (a large model), but not for the smaller CBR models.

目 的

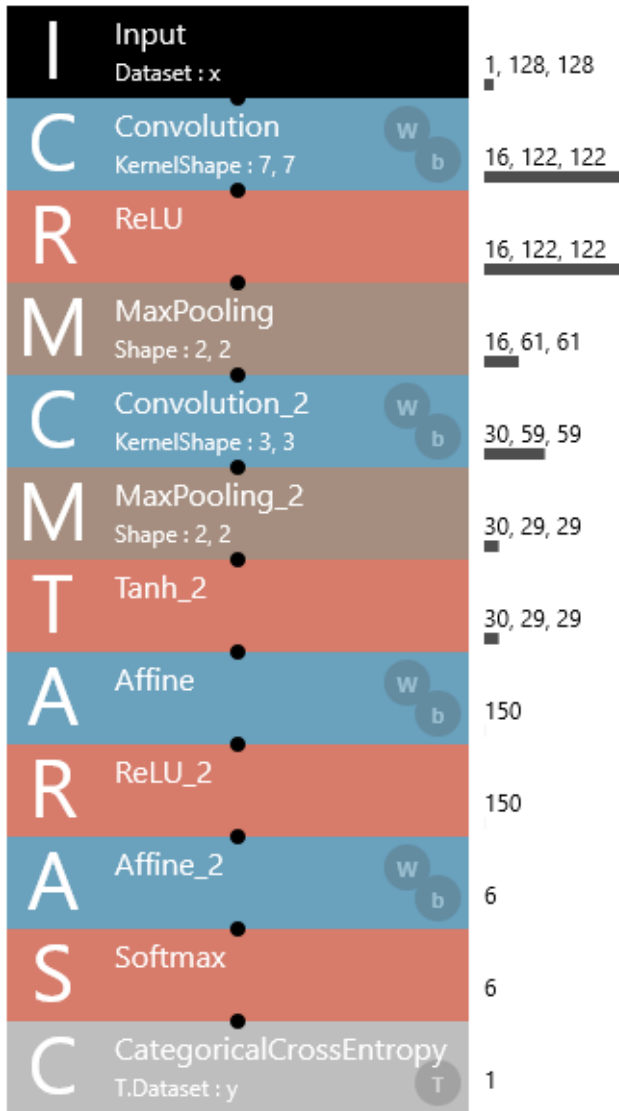
- 【実験1】: 10枚法のデンタルのマウント位置を判定させる目的のため、重厚で複雑なシステムではなく、0から9までの手書き文字(MNISTデータ)を認識する程度の機械学習システムが利用可能かどうか、画像の解像度の影響はあるのか、を明らかにする。
- 【実験2】: 人間が特徴量として捉えている解剖構造を認識しているかを明らかにする。

使用機器、ソフトウェア

- ハードウェア
 - CPU: Intel Core i5-7400 (3GHz)、メモリ: 24GB
 - GPU: NVIDIA GeForce GTX 1060 6GB
 - ソフトウェア (MNISTデータを扱えるレベル)
 - Neural Network Console (ソニー社製、version 1.4.6983.42039)
- LeNet
 - Binary_connect_mnist_LeNet
 - 12_residual_learning (ResNetの亜系)

LeNet

入力層ピクセルサイズと出力層のカテゴリ数のみ変更



入力層

畳込み層

活性化関数 (ReLU: Rectified Linear Unit)

プーリング層 (最大値)

畳込み層

プーリング層 (最大値)

活性化関数 (Tanh)

全結合層

活性化関数 (ReLU)

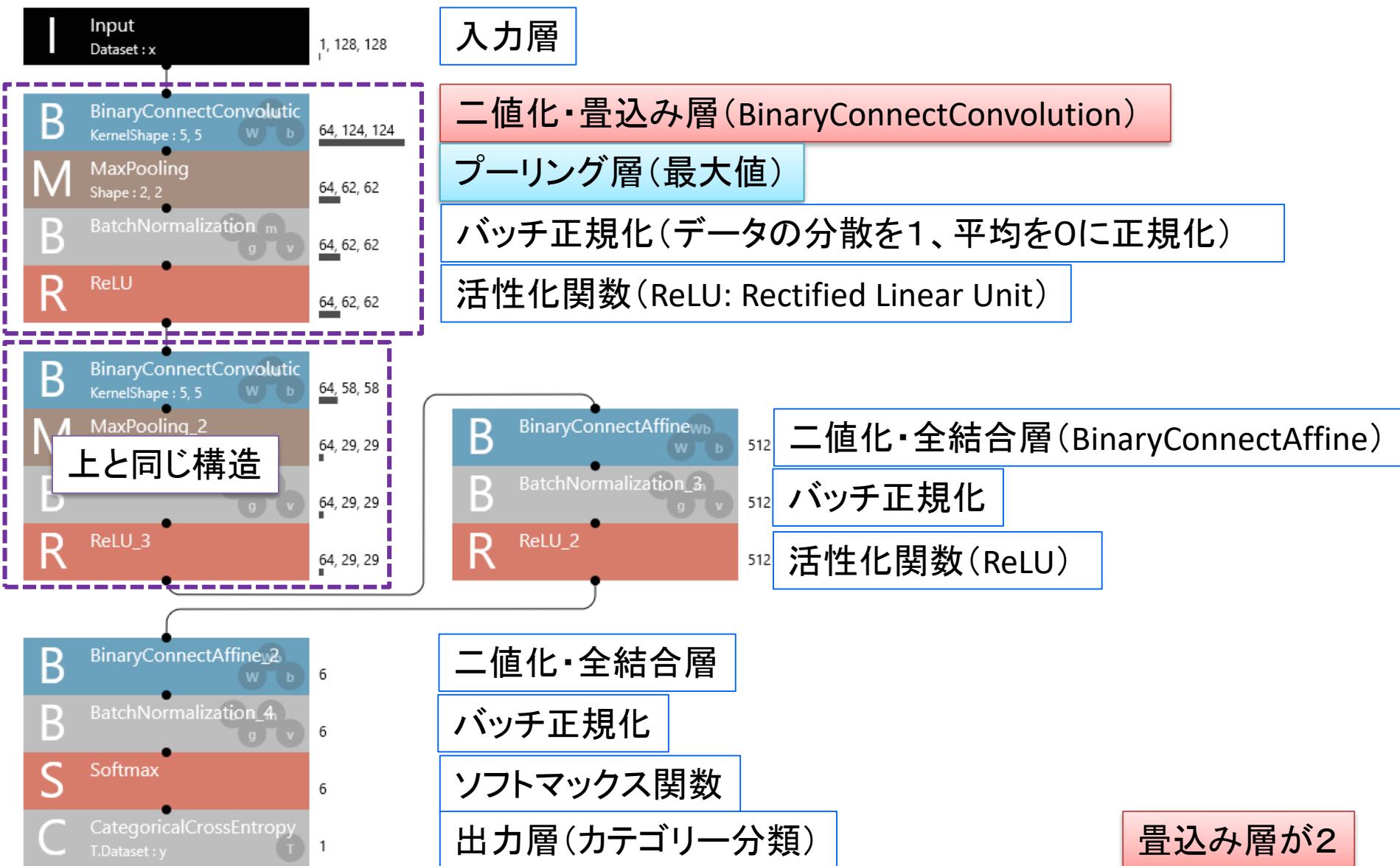
全結合層

ソフトマックス関数

出力層 (カテゴリ分類)

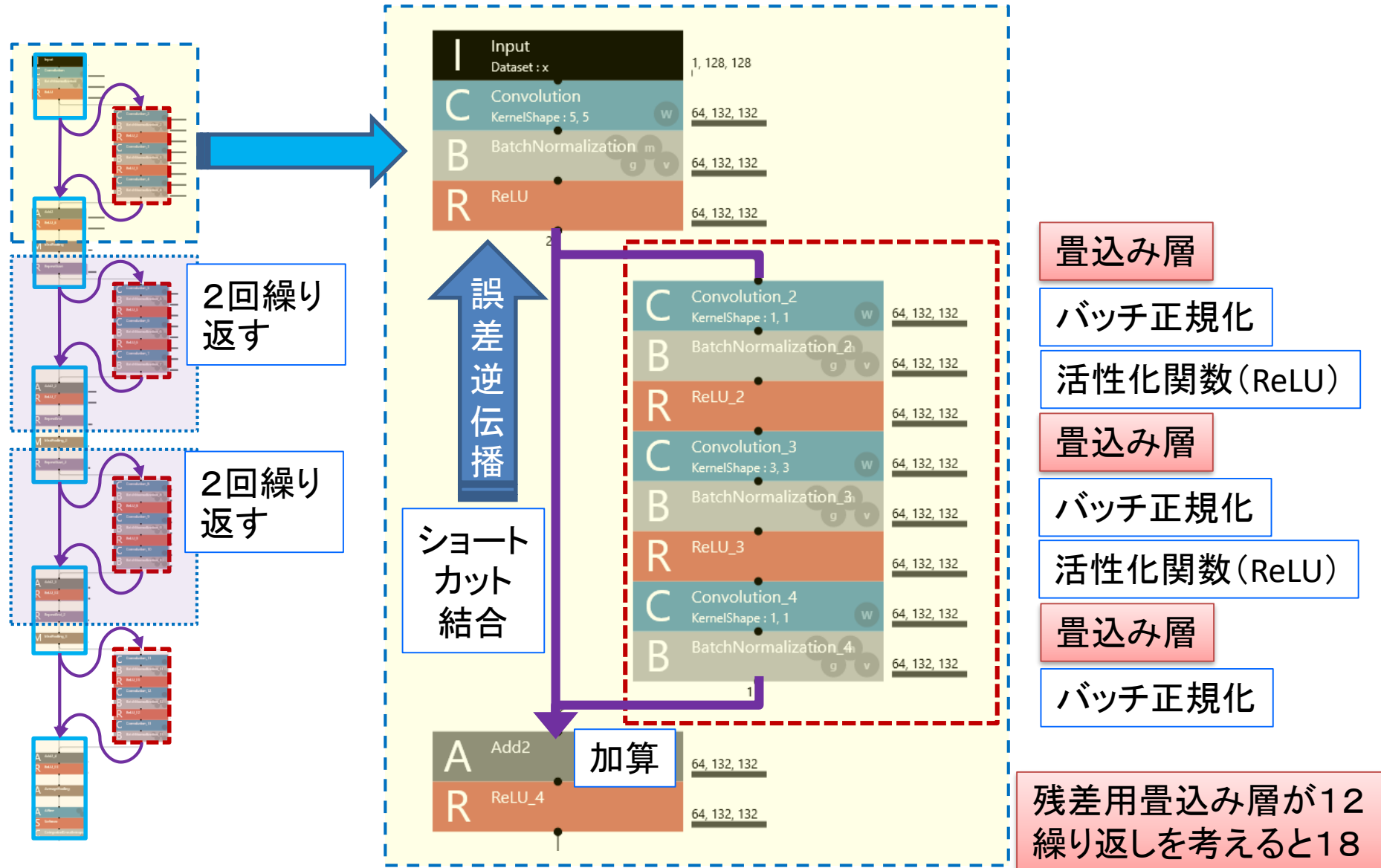
畳み込み層が2

Binary_connect_mnist_LeNet（以後、BC-LeNet） 入力層ピクセルサイズと出力層のカテゴリ数のみ変更



12_residual_learning (以後、mResNet-12)

入力層ピクセルサイズと出力層のカテゴリ数のみ変更

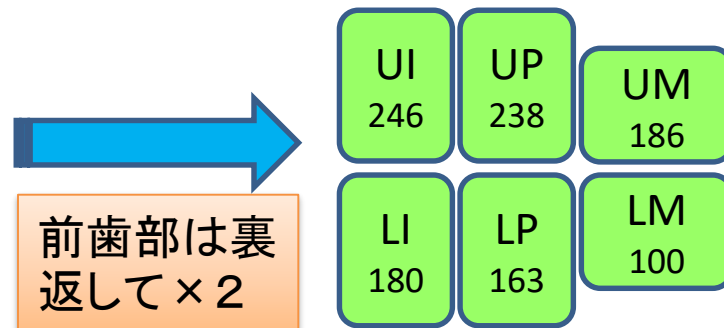
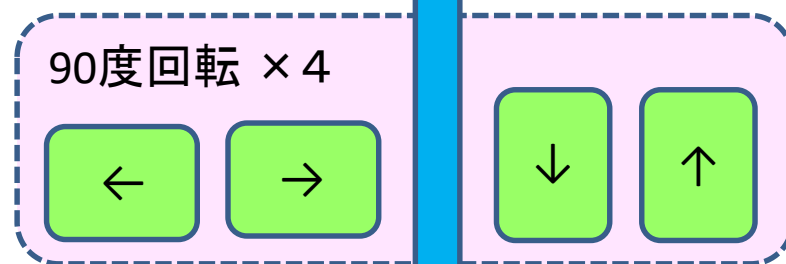
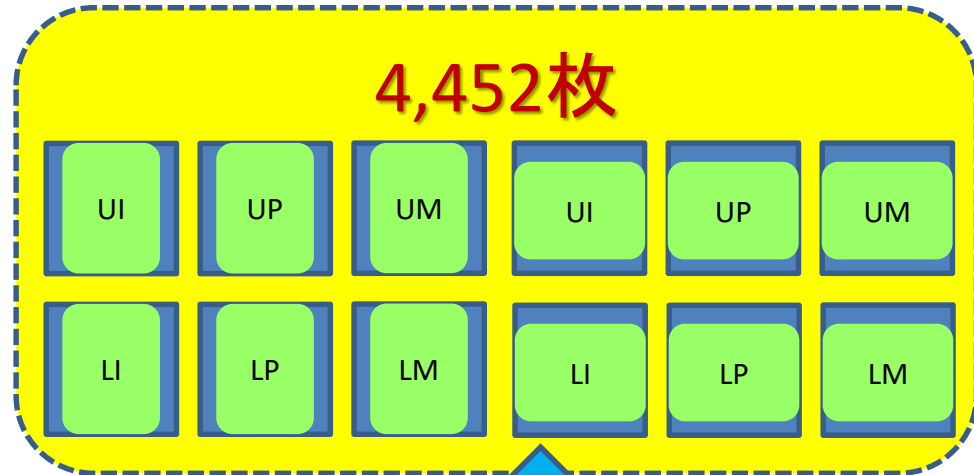
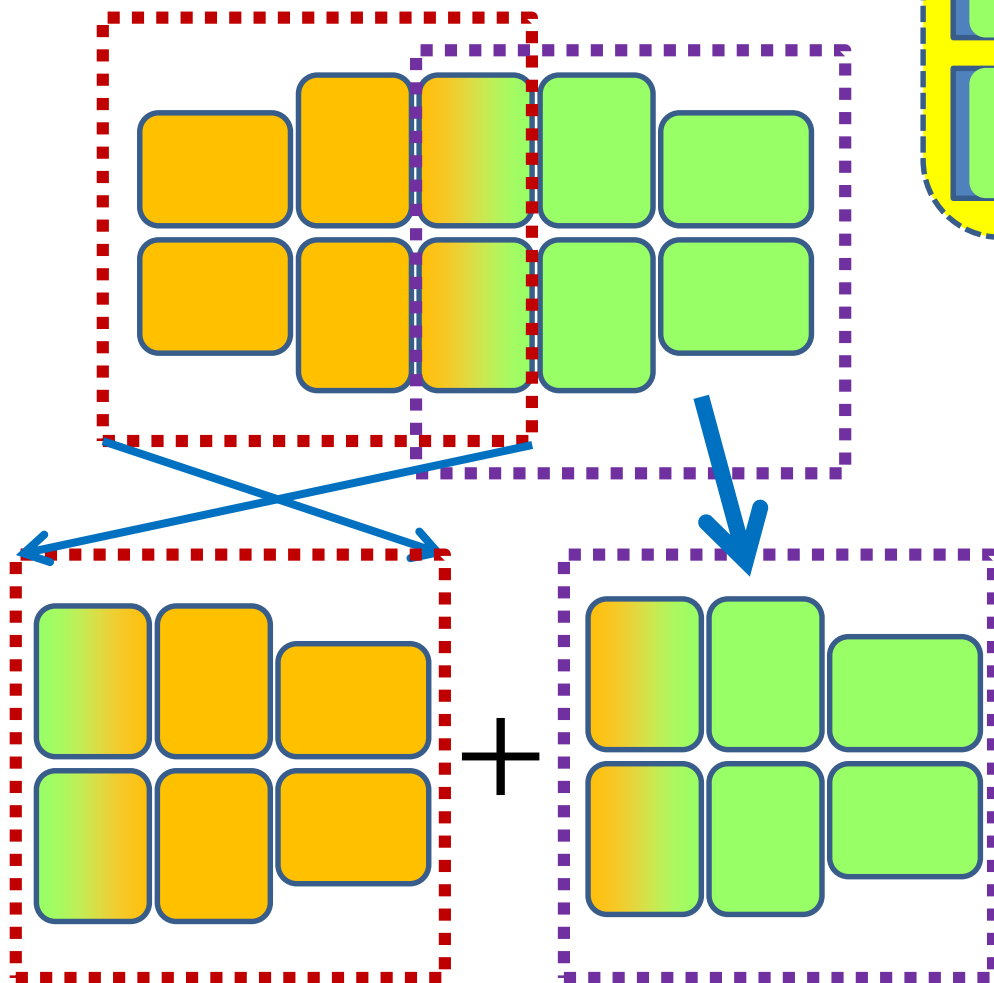


実験1 対象 学習データ

- 10枚法の口内法エックス線画像**150組**、計**1,417枚**の内、明らかな失敗やフィルム内にて智歯を除く1歯以上の欠損やインプラントを除いた**900枚**。
- 画像の右側を反転させることで左側6ブロックとし、90度単位で回転させた**4,452枚**を作成。

対象画像データの生成

150組、計1,417枚から選択した900枚(欠損、明らかな失敗、インプラント体等除く)



前歯部は裏返して×2

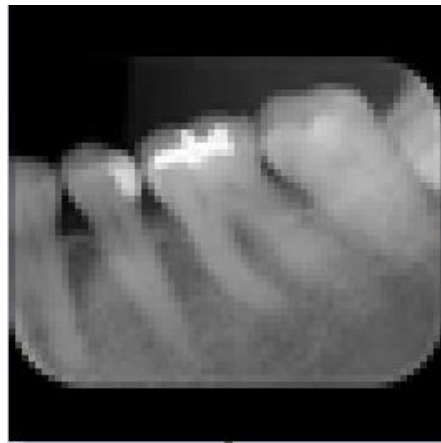
学習データ、検証データ、 テストデータおよび評価方法

- **テスト用**として各部位48枚、計**288枚**を確保。残り4,164枚の内、**学習用**として**80%の3,331枚**を、**学習状態チェック用の検証データ**として**20%の833枚**を用いた。
- 各データセットに**背景色を追加**して方形とし、一辺が**32、64、および128ピクセルの3種類**を生成した。
- 対象画像は、**比較的均質で選択バイアスの強い画像**（欠損歯数の多い画像、明らかな失敗、ないしインプラント等を除いた画像）であるため、クロスバリデーションは用いず、**ホールドアウト法**を用いることとした。

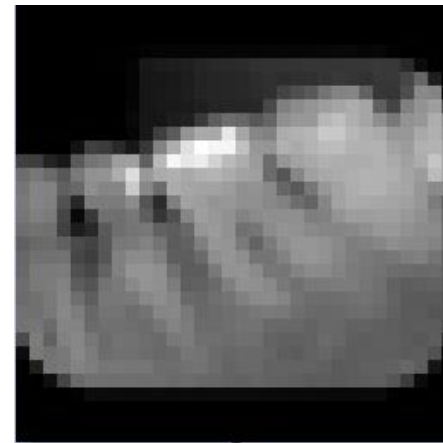
画像データおよび略号



128x128



64x64

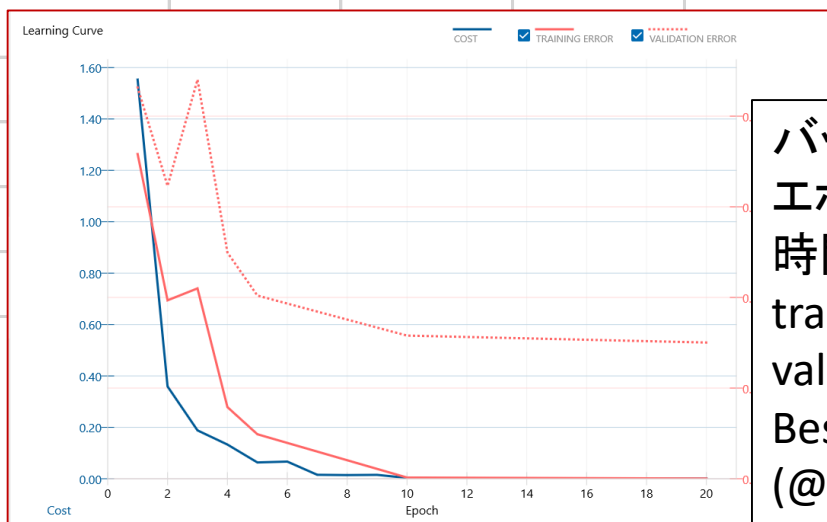


32x32

記号	内容
UI	上顎前歯部
UP	上顎犬歯・小臼歯部
UM	上顎大臼歯部
LI	下顎前歯部
LP	下顎犬歯・小臼歯部
LM	下顎大臼歯部

LeNet 128x128

		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	47	1	0	0	0	0.979
	UM	0	1	43	0	0	4	0.896
	LI	1	0	0	44	3	0	0.917
	LP	0	0	0	1	46	1	0.958
	LM	0	1	6	0	0	41	0.854
Precision		0.980	0.959	0.860	0.978	0.939	0.891	
F-Measures		0.990	0.969	0.878	0.946	0.948	0.872	
Accuracy		0.934						
Avg.Precision		0.934						
Avg.Recall		0.934						
Avg.F-Measures		0.934						

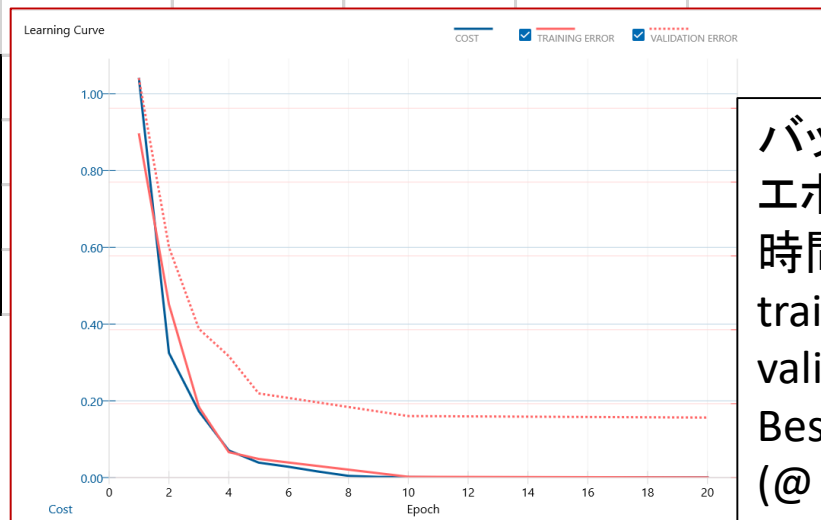


バッチ数: 20
 エポック数: 20
 時間 00:00:49
 train_error=0.000205
 valid_error=0.150
 Best valid_error=0.150
 (@ epoch 20)

LeNet 64x64

		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	1	47	0	0	0	0.979
	LI	0	0	0	47	1	0	0.979
	LP	0	0	0	0	48	0	1.000
	LM	0	0	1	0	0	47	0.979
Precision		1.000	0.980	0.979	1.000	0.980	1.000	
F-Measures		1.000	0.990	0.979	0.989	0.990	0.989	

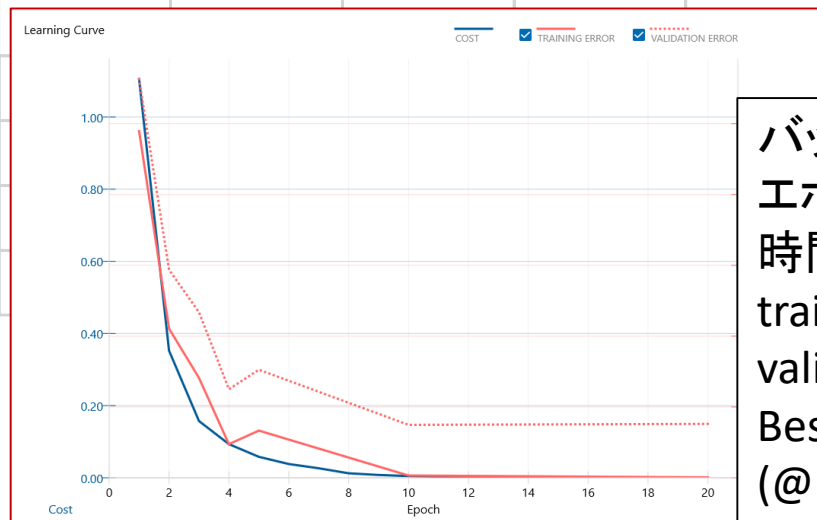
Accuracy	0.990
Avg.Precision	0.990
Avg.Recall	0.990
Avg.F-Measures	0.990



バッチ数: 20
 エポック数: 20
 時間 00:00:30
 train_error=0.000200
 valid_error=0.0812
 Best valid_error= 0.0812
 (@ epoch 20)

LeNet 32x32

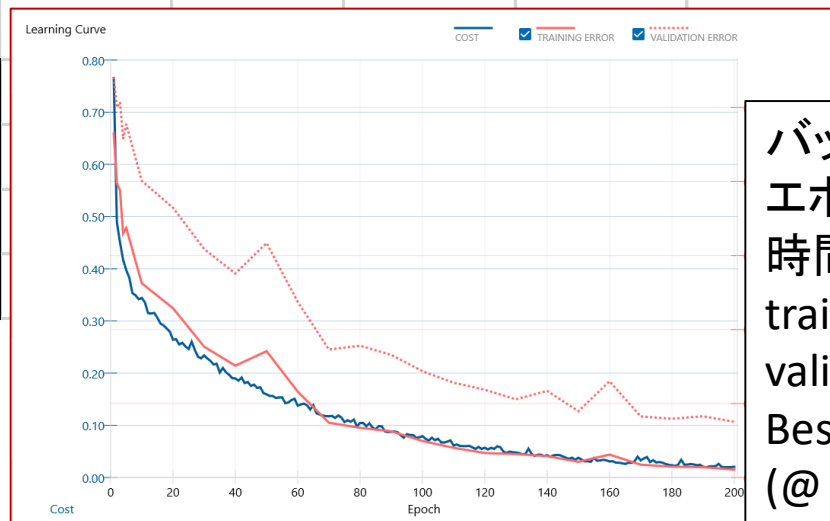
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	47	0	0	1	0.979
	LI	0	0	0	48	0	0	1.000
	LP	0	0	1	0	46	1	0.958
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	0.979	1.000	1.000	0.960	
F-Measures		1.000	1.000	0.979	1.000	0.979	0.980	
Accuracy		0.990						
Avg.Precision		0.990						
Avg.Recall		0.990						
Avg.F-Measures		0.990						



バッチ数: 20
 エポック数: 20
 時間 00:00:20
 train_error=0.000530
 valid_error=0.0760
 Best valid_error= 0.0748
 (@ epoch 10)

BC-LeNet 128x128

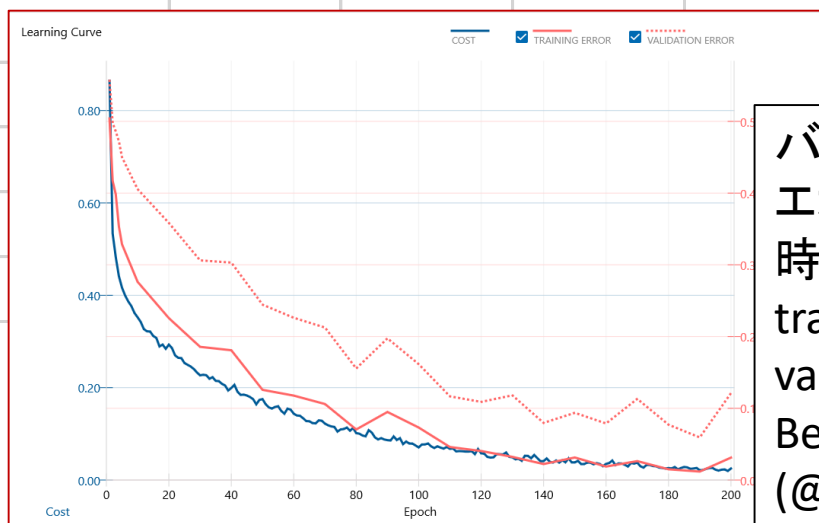
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	46	2	0	0	0	0.958
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	45	3	0	0.938
	LP	0	0	0	0	48	0	1.000
	LM	0	0	3	0	0	45	0.938
Precision		1.000	1.000	0.906	1.000	0.941	1.000	
F-Measures		1.000	0.979	0.950	0.968	0.970	0.968	
Accuracy		0.972						
Avg.Precision		0.974						
Avg.Recall		0.972						
Avg.F-Measures		0.972						



バッチ数: 30
 エポック数: 200
 時間 00:23:03
 train_error=0.0107
 valid_error=0.0754
 Best valid_error=0.0754
 (@ epoch 200)

BC-LeNet 64x64

		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	1	0	0	47	0.979
Precision		1.000	1.000	0.980	1.000	1.000	1.000	
F-Measures		1.000	1.000	0.990	1.000	1.000	0.989	
Accuracy		0.997						
Avg.Precision		0.997						
Avg.Recall		0.997						
Avg.F-Measures		0.997						

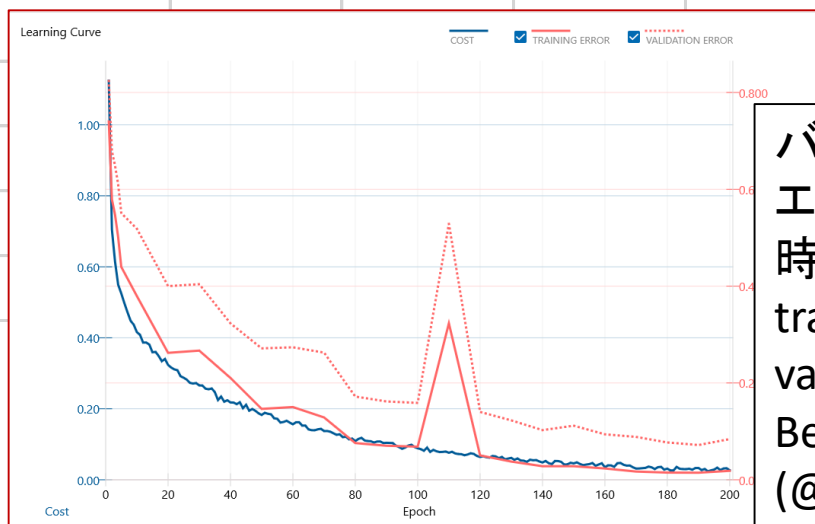


バッチ数: 30
 エポック数: 200
 時間 00:05:16
 train_error=0.0312
 valid_error=0.121
 Best valid_error=0.0594
 (@ epoch 190)

BC-LeNet 32x32

		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	1	0	0	47	0.979
Precision		1.000	1.000	0.980	1.000	1.000	1.000	
F-Measures		1.000	1.000	0.990	1.000	1.000	0.989	

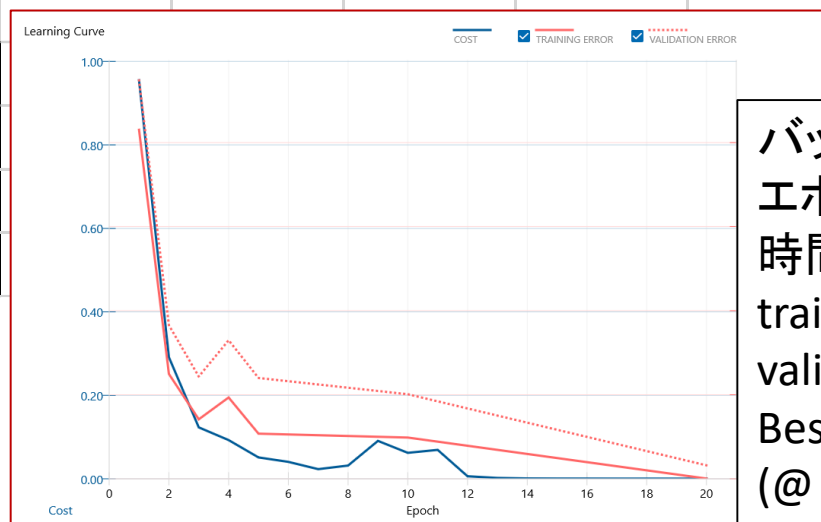
Accuracy	0.997
Avg.Precision	0.997
Avg.Recall	0.997
Avg.F-Measures	0.997



バッチ数: 30
 エポック数: 200
 時間 00:02:00
 train_error=0.0185
 valid_error=0.0841
 Best valid_error=0.0719
 (@ epoch 190)

mResNet-12 128x128

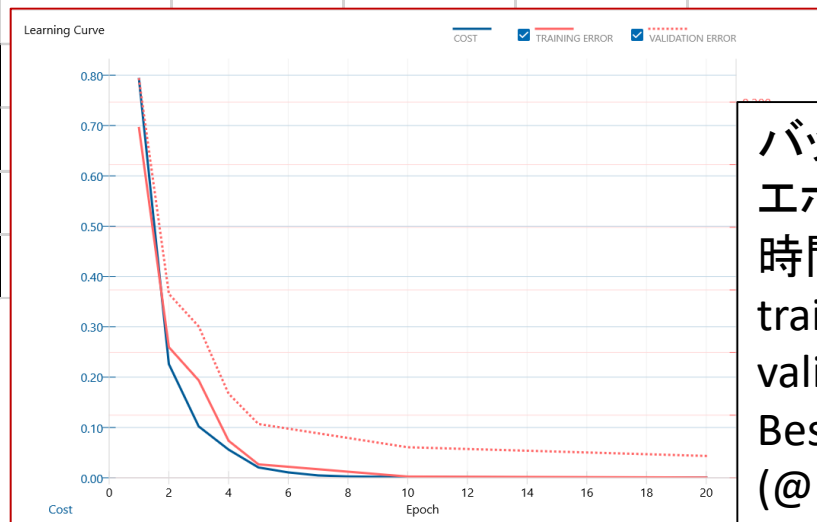
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	1.000	1.000	1.000	1.000	
F-Measures		1.000	1.000	1.000	1.000	1.000	1.000	
Accuracy		1.000						
Avg.Precision		1.000						
Avg.Recall		1.000						
Avg.F-Measures		1.000						



バッチ数: 20
 エポック数: 20
 時間 00:07:49
 train_error=0.000277
 valid_error=0.0161
 Best valid_error= 0.0161
 (@ epoch 20)

mResNet-12 64x64

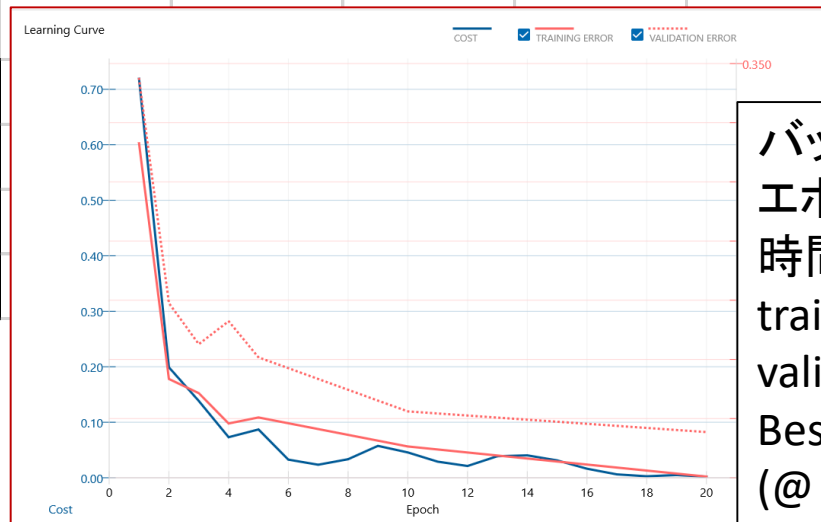
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	1.000	1.000	1.000	1.000	
F-Measures		1.000	1.000	1.000	1.000	1.000	1.000	
Accuracy		1.000						
Avg.Precision		1.000						
Avg.Recall		1.000						
Avg.F-Measures		1.000						



バッチ数: 20
 エポック数: 20
 時間 00:02:18
 train_error=0.000198
 valid_error=0.0175
 Best valid_error= 0.0175
 (@ epoch 20)

mResNet-12 32x32

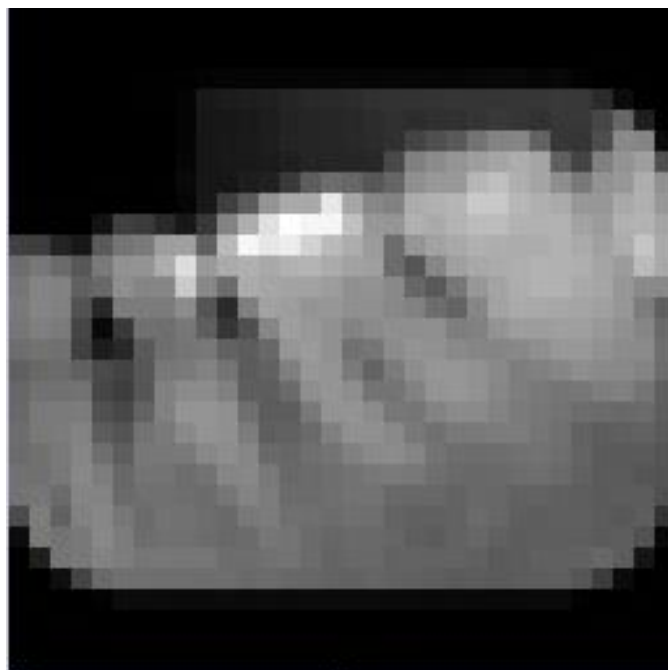
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	47	0	0	1	0.979
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	1.000	1.000	1.000	0.980	
F-Measures		1.000	1.000	0.989	1.000	1.000	0.990	
Accuracy		0.997						
Avg.Precision		0.997						
Avg.Recall		0.997						
Avg.F-Measures		0.997						



バッチ数: 20
 エポック数: 20
 時間 00:00:52
 train_error=0.000973
 valid_error=0.0386
 Best valid_error= 0.0386
 (@ epoch 20)

32ピクセル画像の人間による判定

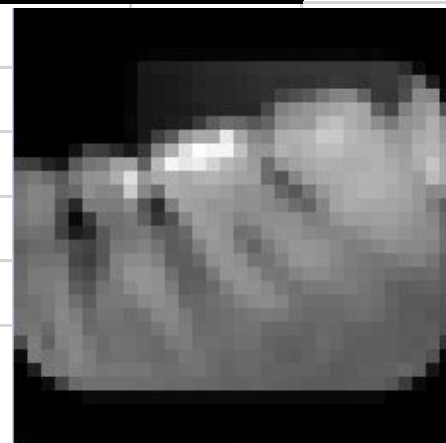
- テストデータの6部位から3枚ずつランダムに選択した32ピクセル画像18枚。人間(9名の共同演者)とAIにて評価し、比較した。



32x32 人間での評価

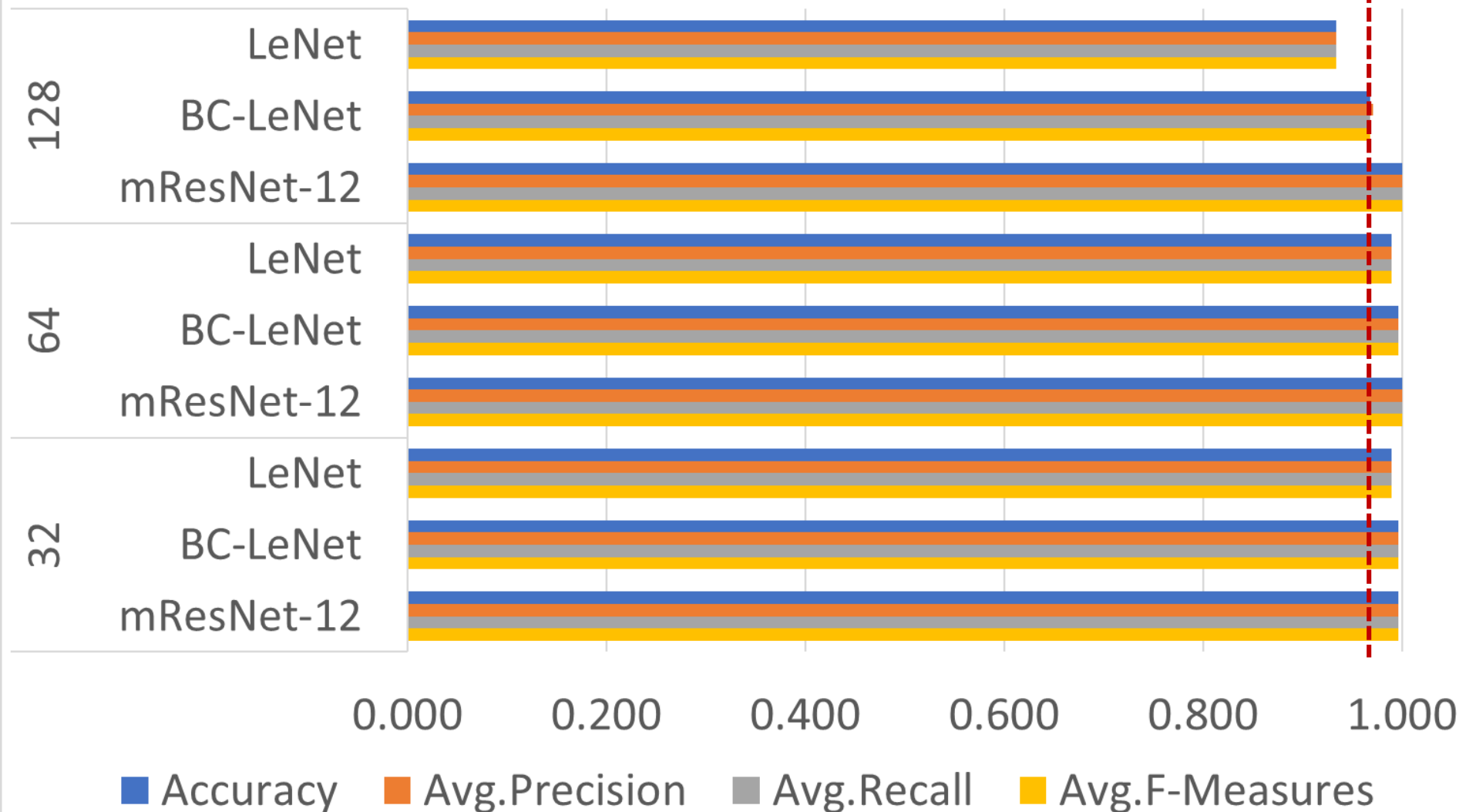
		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	27	0	0	0	0	0	1.000
	UP	0	26	0	0	1	0	0.963
	UM	0	1	25	0	0	1	0.926
	LI	0	0	0	27	0	0	1.000
	LP	0	2	0	0	25	0	0.926
	LM	0	0	0	0	0	27	1.000
Precision		1.000	0.897	1.000	1.000	0.962	0.964	
F-Measures		1.000	0.929	0.981	1.000	0.943	0.982	
Accuracy		0.969						
Avg.Precision		0.970						
Avg.Recall		0.969						
Avg.F-Measures		0.972						

約97%



人間による32x32
画像の評価

実験 1 ・ 学習結果



実験1 結果・まとめ

- デンタルのマウント位置の判定には、0から9までの手書き文字(MNISTデータ)を認識する程度の機械学習システムが利用可能であり、有能であることが分かった。
- 畳込み層の数が少ないと、画像の解像度が高い場合に診断精度が劣化する傾向がみられた。ショートカットを入れた12層(繰り返し込みで18層)の残差ネットの構造を有するシステムでは、診断精度にほとんど変化が見られなかった。

実験2

- 実験1は大成功だったといえるが、ブラックボックスを評価するにはどうすればよい？
- ヒートマップや勾配ベースハイライト法等があるが、画像診断にて人間が理解する指標といえは解剖構造だろう。
- したがって、教科書にも使用されているシェーマを使うこととした。
- 一番成績の良かったmResNet-12のみ、結果を提示する。

実験2 対象 評価用データ

- **【濃淡シェーマデータ】**: 学生提示用のシェーマにて、できる限り濃淡を実際の画像に合わせるようにした6部位6枚。
- **【線画シェーマデータ】**: 2冊の教科書(※)から線画のシェーマを抜き出した6部位15枚。
- シェーマデータは、各部位ランダムにコピーを生成し、1部位毎に8枚となるようにし、ランダムに回転させた。

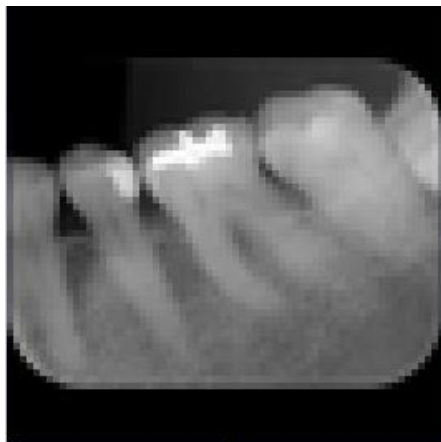
※歯科衛生士テキスト、わかりやすい歯科放射線学(第2版)、学建書院

※歯科放射線学、第6版、医歯薬出版

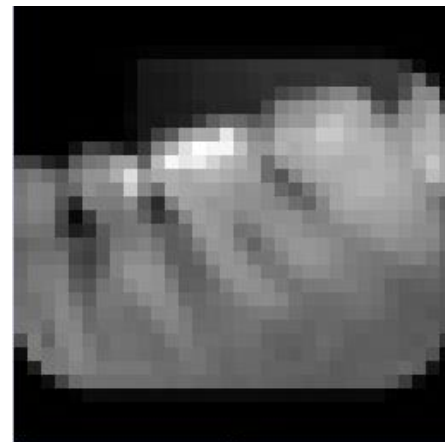
画像データとシェーマおよび略号



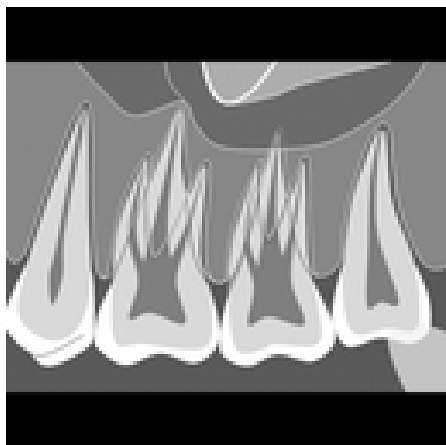
128x128



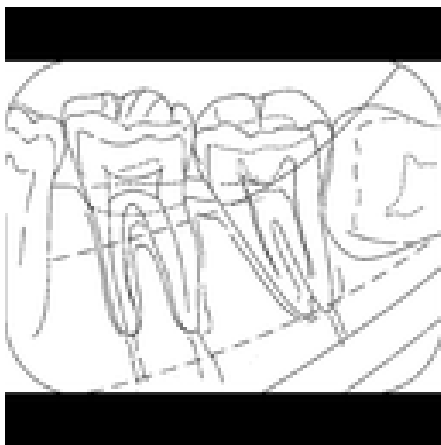
64x64



32x32



濃淡シェーマ



線画シェーマ

記号	内容
UI	上顎前歯部
UP	上顎犬歯・小臼歯部
UM	上顎大臼歯部
LI	下顎前歯部
LP	下顎犬歯・小臼歯部
LM	下顎大臼歯部

128 x 128

予測した部位

UI	UP	UM	LI	LP	LM	Recall
0	0	0	1	0	4	0.000
0	0	0	1	0	4	0.000
0	0	2	2	0	1	0.400
0	0	0	2	0	3	0.400
0	0	0	5	0	0	0.000
0	0	0	0	0	5	1.000
0.000	0.000	1.000	0.182	0.000	0.294	
0.000	0.000	0.571	0.250	0.000	0.455	

LeNet

Accuracy

17 - 30 % 程度



64 x 64

予測した部位

	64 x 64	予測された部位						
		UI	UP	UM	LI	LP	LM	Recall
真の 部位	UI	0	0	1	0	4	0	0.000
	UP	0	0	0	0	5	0	0.000
	UM	0	0	0	2	3	0	0.000
	LI	0	0	0	1	4	0	0.200
	LP	0	0	0	0	5	0	1.000
	LM	0	0	0	4	1	0	0.000
	Precision	0.000	0.000	0.000	0.143	0.227	0.000	
	F-Measures	0.000	0.000	0.000	0.167	0.370	0.000	

Accuracy

0.200

Avg.Precision

0.062

Avg.Recall

0.200

Avg.F-Measures

0.089

32 x 32

予測した部位

	32 x 32	UI	UP	UM	LI	LP	LM	Recall
真の 部 位	UI	0	0	1	4	0	0	0.000
	UP	0	0	0	5	0	0	0.000
	UM	0	0	0	4	1	0	0.000
	LI	0	0	0	5	0	0	1.000
	LP	0	1	0	4	0	0	0.000
	LM	0	0	0	4	1	0	0.000
	Precision	0.000	0.000	0.000	0.192	0.000	0.000	
	F-Measures	0.000	0.000	0.000	0.323	0.000	0.000	

Accuracy

0.167

Avg.Precision

0.032

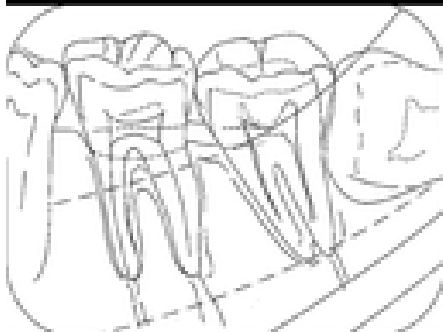
Avg.Recall

0.167

Avg.F-Measures

0.054

線画シェーマ



128 x 128

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	0	0	5	0	0.000
	UP	0	0	0	5	0	0.000
	UM	0	0	0	5	0	0.000
	LI	0	0	0	5	0	1.000
	LP	0	0	0	5	0	0.000
	LM	0	0	0	5	0	0.000
Precision	0.000	0.000	0.000	0.167	0.000	0.000	
F-Measures	0.000	0.000	0.000	0.286	0.000	0.000	

BC-LeNet

Accuracy

17 - 23 % 程度



64 x 64

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	0	0	5	0	0.000
	UP	0	0	0	5	0	0.000
	UM	0	0	1	4	0	0.200
	LI	0	0	0	5	0	1.000
	LP	0	0	0	5	0	0.000
	LM	0	0	0	5	0	0.000
Precision	0.000	0.000	1.000	0.172	0.000	0.000	
F-Measures	0.000	0.000	0.333	0.294	0.000	0.000	

Accuracy	0.200
Avg.Precision	0.195
Avg.Recall	0.200
Avg.F-Measures	0.105

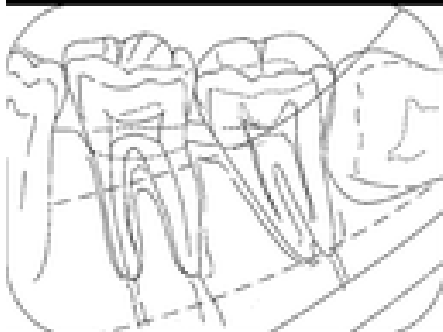
32 x 32

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	1	0	0	4	0	0.200
	UP	0	0	0	4	1	0.000
	UM	0	0	0	2	3	0.000
	LI	0	0	0	2	3	0.400
	LP	1	0	0	0	4	0.800
	LM	3	0	0	0	2	0.000
Precision	0.200	0.000	0.000	0.167	0.308	0.000	
F-Measures	0.200	0.000	0.000	0.235	0.444	0.000	

Accuracy	0.233
Avg.Precision	0.112
Avg.Recall	0.233
Avg.F-Measures	0.147

線画シェーマ



128 x 128

予測した部位

UI	UP	UM	LI	LP	LM	Recall
0	5	0	0	0	0	0.000
0	5	0	0	0	0	1.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
Precision	0.000	0.167	0.000	0.000	0.000	
F-Measures	0.000	0.286	0.000	0.000	0.000	

mResNet-12

Accuracy

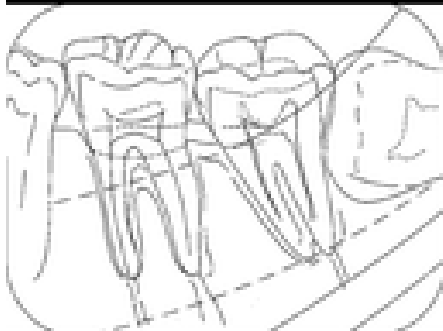
17 - 20 % 程度

64 x 64

予測した部位

UI	UP	UM	LI	LP	LM	Recall
0	5	0	0	0	0	0.000
0	5	0	0	0	0	1.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
0	5	0	0	0	0	0.000
Precision	0.000	0.167	0.000	0.000	0.000	
F-Measures	0.000	0.286	0.000	0.000	0.000	

線画シェーマ



32 x 32

予測した部位

UI	UP	UM	LI	LP	LM	Recall
5	0	0	0	0	0	1.000
5	0	0	0	0	0	0.000
5	0	0	0	0	0	0.000
4	0	0	1	0	0	0.200
1	0	0	4	0	0	0.000
4	0	0	1	0	0	0.000
Precision	0.208	0.000	0.000	0.167	0.000	
F-Measures	0.345	0.000	0.000	0.182	0.000	
Accuracy	0.200					
Avg.Precision	0.063					
Avg.Recall	0.200					
Avg.F-Measures	0.088					

128 x 128

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	4	0	0	2	0.000
	UP	0	4	2	0	2	0.500
	UM	0	0	4	0	4	0.500
	LI	0	0	0	6	2	0.750
	LP	0	4	0	0	2	0.250
	LM	0	0	0	0	8	1.000
Precision	0.000	0.333	0.667	1.000	0.500	0.400	
F-Measures	0.000	0.400	0.571	0.857	0.333	0.571	

LeNet

Accuracy

17 - 50 % 程度



64 x 64

予測した部位

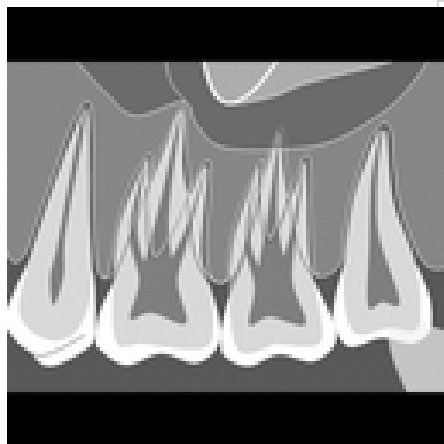
	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	4	2	0	0	0.000
	UP	2	6	0	0	0	0.750
	UM	0	0	6	0	0	0.750
	LI	0	8	0	0	0	0.000
	LP	0	8	0	0	0	0.000
	LM	0	0	8	0	0	0.000
Precision	0.000	0.231	0.375	0.000	0.000	0.000	
F-Measures	0.000	0.353	0.500	0.000	0.000	0.000	

32 x 32

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	2	0	6	0	0	0.250
	UP	8	0	0	0	0	0.000
	UM	0	0	4	0	0	0.500
	LI	6	2	0	0	0	0.000
	LP	0	8	0	0	0	0.000
	LM	0	0	6	0	0	0.250
Precision	0.125	0.000	0.250	0.000	0.000	0.333	
F-Measures	0.167	0.000	0.333	0.000	0.000	0.286	

濃淡シェーマ



Accuracy	0.167
Avg.Precision	0.118
Avg.Recall	0.167
Avg.F-Measures	0.131

128 x 128

予測した部位

UI	UP	UM	LI	LP	LM	Recall
0	0	4	4	0	0	0.000
0	0	0	8	0	0	0.000
0	0	8	0	0	0	1.000
0	0	0	8	0	0	1.000
2	4	0	2	0	0	0.000
0	0	6	0	0	2	0.250
0.000	0.000	0.444	0.364	0.000	1.000	
0.000	0.000	0.615	0.533	0.000	0.400	

BC-LeNet

Accuracy

33 - 42 % 程度



64 x 64

予測した部位

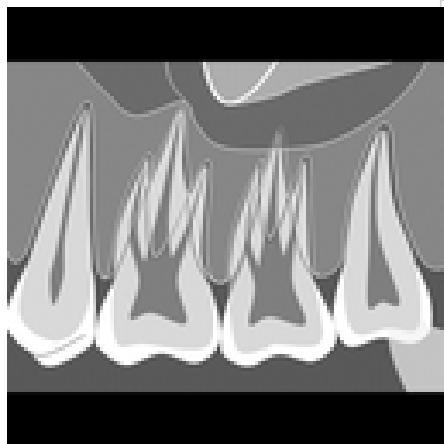
	64 x 64	予測された部位						
		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	6	0	2	0	0	0	0.750
	UP	6	0	0	2	0	0	0.000
	UM	0	0	6	0	0	2	0.750
	LI	0	2	6	0	0	0	0.000
	LP	2	0	6	0	0	0	0.000
	LM	0	0	4	0	0	4	0.500
	Precision	0.429	0.000	0.250	0.000	0.000	0.667	
	F-Measures	0.545	0.000	0.375	0.000	0.000	0.571	

32 x 32

予測した部位

Accuracy	0.333	真 の 部 位	32 x 32						Recall
Avg.Precision	0.224		UI	8	0	0	0	0	1.000
Avg.Recall	0.333		UP	8	0	0	0	0	0.000
Avg.F-Measures	0.249		UM	0	0	8	0	0	1.000
			LI	4	2	0	2	0	0.250
			LP	2	4	2	0	0	0.000
			LM	0	0	6	0	0	0.250
			Precision	0.364	0.000	0.500	1.000	0.000	1.000
		F-Measures	0.533	0.000	0.667	0.400	0.000	0.400	

濃淡シェーマ



Accuracy	0.417
Avg.Precision	0.477
Avg.Recall	0.417
Avg.F-Measures	0.333

128 x 128

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	2	2	0	0	4	0.250
	UP	2	4	2	0	0	0.500
	UM	4	0	4	0	0	0.500
	LI	8	0	0	0	0	0.000
	LP	0	4	0	4	0	0.500
	LM	0	0	6	0	2	0.250
Precision		0.125	0.400	0.333	0.000	1.000	0.333
F-Measures		0.167	0.444	0.400	0.000	0.667	0.286

mResNet-12

Accuracy

25 - 33 % 程度

64 x 64

予測した部位

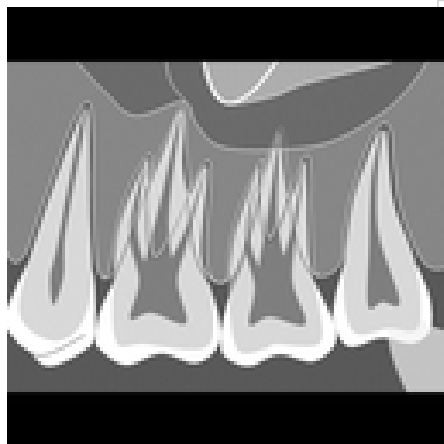
	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	0	2	0	6	0.000
	UP	0	0	0	0	8	0.000
	UM	0	0	8	0	0	1.000
	LI	2	2	0	0	4	0.000
	LP	0	0	0	0	8	0.000
	LM	2	0	2	0	4	0.500
Precision		0.000	0.000	0.667	0.000	0.000	0.133
F-Measures		0.000	0.000	0.800	0.000	0.000	0.210

32 x 32

予測した部位

	UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	6	0	0	0	2	0.750
	UP	6	2	0	0	0	0.250
	UM	2	0	6	0	0	0.750
	LI	4	4	0	0	0	0.000
	LP	4	4	0	0	0	0.000
	LM	6	0	0	0	2	0.250
Precision		0.214	0.200	1.000	0.000	0.000	0.500
F-Measures		0.333	0.222	0.857	0.000	0.000	0.333

濃淡シェーマ



Accuracy	0.333
Avg.Precision	0.319
Avg.Recall	0.333
Avg.F-Measures	0.291

実験2 結果

- 線画シェーマの判別は、全てのシステムで accuracy は最大で30%程度止まりで、判別できなかった。
- 濃淡シェーマの判別は、線画シェーマに比べて良好だったが、accuracy は最大でも50%止まりであり、こちらも判別できなかった。

まとめ

- 0から9までの手書き文字(MNISTデータ)を認識するシステムにて、6ブロックのみでの検証であるが、10枚法のマウント位置決定が高精度で可能なことが分かった。
- デンタルマウントの位置を機械学習システムで判別させる場合、人間が特徴量として捉えている解剖構造は、ほとんど利用されていないと思われた。
- **学生への教育時には、機械学習の有用性とともに、システムによっては判断根拠は人間とは異なりうることを教える必要があると思われた。**