

2021 年度卒業

修士論文

空間解像度差のあるデータセットを用いた深層学習による銀河形状分類
精度

日本語タイトル暫定版のため，ここ英語タイトルも未完！

所属	新潟大学自然科学研究科 電気情報工学専攻・飯田佑輔研究室
在籍番号	F20C026D
氏名	本間 裕也

概要

日本語のアブストラクト

Abstract

English Abstract Here

目次

第1章	はじめに	1
第2章	深層学習	2
2.1	パーセプトロン	2
2.2	ニューラルネットワーク	2
2.3	損失関数と重み更新	3
第3章	使用するデータセット	4
3.1	Sloan Digital Sky Survey(SDSS)	4
3.2	Galaxy Zoo (GZ)	6
第4章	SDSS & Galaxy Zoo を用いた分類モデル学習	8
4.1	実験概要	8
4.2	不確かラベルが与える分類精度への影響	12
4.2.1	実験条件	13
4.2.2	実験結果	13
4.3	データインバランスが与える分類精度への影響	17
4.3.1	実験条件	17
4.3.2	実験結果	18
4.4	4章全体の議論・結論	21
第5章	空間解像度差のあるデータセットを用いた分類モデル学習	23
5.1	実験概要	23
5.2	分類精度のデータセット内解像度差依存性	27
5.2.1	実験条件	27
5.2.2	実験結果	28
5.3	分類精度の使用画像枚数依存性	34
5.3.1	実験条件	34

5.3.2	実験結果	35
5.4	5 章全体の議論・結論	39
第 6 章	議論	43
6.1	今回の実験から得た結論	43
6.2	将来課題	43
6.3	将来展望	44
6.3.1	実験概要	44
6.3.2	実験結果	46
6.3.3	結果	46
第 7 章	おわりに	47

第1章 はじめに

第2章 深層学習

2.1 パーセプトロン

深層学習の説明を行う前に,

2.2 ニューラルネットワーク

ニューラルネットワークとは,

2.3 損失関数と重み更新

深層学習の学習で用いられる指標を、損失関数と呼ぶ。損失関数には様々な種類が存在し、解く問題の種類によって使い分ける。一般的な損失関数として、式 (2.1) の 2 乗平均誤差 (主に回帰問題に使用) や、式 (2.2) のクロスエントロピー誤差 (主に分類問題に使用) が挙げられる。

$$E = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2.1)$$

$$E = \text{cross entropy} \quad (2.2)$$

第3章 使用するデータセット

3.1 Sloan Digital Sky Survey(SDSS)

Sloan Digital Sky Survey (以下, SDSS) について

SDSSとは、天文学史上最も重要なサーベイ観測プロジェクトの一つとも称される大規模プロジェクトである。このプロジェクトは全天の約1/4の天域において、天体の画像および分光データを取得し、天体カタログを作成することを目的としたものである。SDSSでの撮像・分光データ取得はCCDを搭載した地上望遠鏡によって行われる。SDSSの最初のプロジェクトであるSDSS-Iは2000年から2008年まで行われ、また対象を銀河系や超新星に絞ったSDSS-IIが2005年から2008年にSDSS-Iと並行して実施された。その後、太陽系外惑星調査や天の川銀河の構造及び進化などに焦点を当てたSDSS-IIIが2008年から2014年に、南天・北天からの銀河系探索などを目的としたSDSS-IVが2014年から2020年に行われた。

SDSSで撮影された天体のうち低赤方偏移銀河は後述するGalaxy Zooによって形態分類ラベル付けが行われている。SDSSとGalaxy Zooから天体の画像データと分類ラベルを取得できることから、これらのセットを今回の銀河形態分類モデル作成に用いる。

実験に用いる銀河画像データの作成方法

今回分類モデル学習に用いる銀河画像データとして、SDSS-IIにおけるデータリリースの中から、Data Release 7 (以下, DR7)より画像データの取得を行った。DR7を選んだ理由としては、Galaxy Zooにおける銀河形態ラベル付けにDR7の銀河画像が用いられたからである。

データベースから取得できるのはある程度大きな天域の全体画像のため、用いたい銀河の画像を取得したい場合は、全体画像から切り出しを行う必要がある。今回はGalaxy Zooにて形態ラベル付けが為されている銀河の中から15,000天体を対象に、銀河毎に64ピクセル四方のサイズで切り出しを行った。銀河切り出し画像の生成概略図を図??に示す。

DR7におけるデータの撮影が行われたSDSS-IIにおいて、銀河撮像に用いられた測光フィルタはu, g, r, i, zの5つが存在し、これらのフィルタを使用し5つの帯域画像が撮影された(図3.1参

照). これら 5 つの帯域画像のうち、今回の実験では r フィルタから得られた帯域画像 (r バンド画像) を使用している. r バンド画像を使用した理由としては、5 つの帯域画像のうち可視光帯画像である g, r, i バンド画像の中で、最も平均値に近い波長を捉えている r バンド画像がより多くの銀河形態的特徴を有していると考えられること、また r フィルタが 5 つの測光フィルタのうち最も感度がよいことが挙げられる??.

u	g	r	i	z
3551Å	4686Å	6165Å	7481Å	8931Å
22.0	22.2	22.2	21.3	20.5

図 3.1: SDSS Data Release 7 における、銀河撮像に用いられた測光フィルタ一覧
(フィルタ名、各フィルタによって撮影された画像の波長平均値)

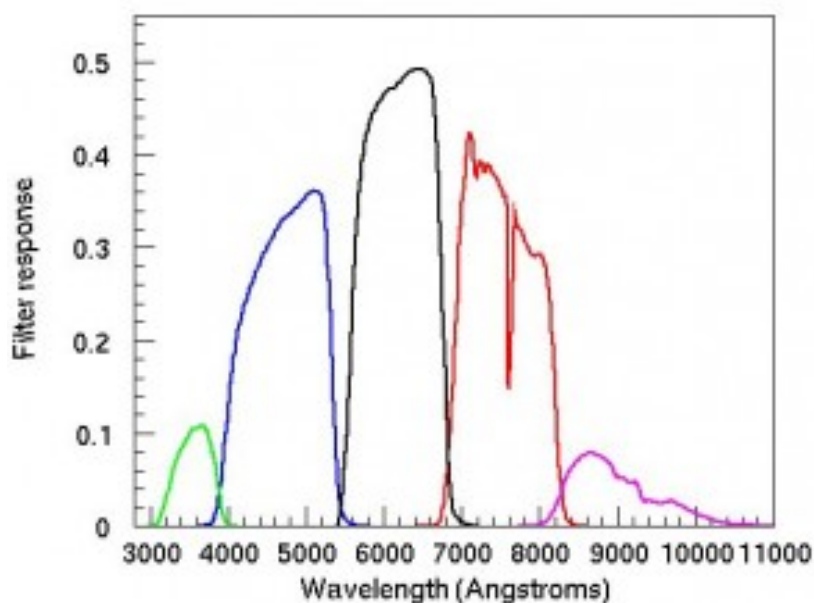


図 3.2: SDSS における測光フィルタのスループット曲線
(<https://www.sdss.org/instruments/camera/> より引用)

3.2 Galaxy Zoo (GZ)

Galaxy Zoo (以下, GZ) について

Galaxy Zoo??とは、人間の目による分類を行い銀河形態カタログを作成したプロジェクトである。従来の銀河形態分類は専門知識を持った天文学者のチームによって行われてきたが、SDSS のような何十万もの銀河が格納されたデータセットが登場するに従い、天文学者だけでは銀河データの増加に追いつけなくなってきた。このような状況を打破する方法として、インターネットを通じて専門知識を持たない有志の一般人に投票形式で形態分類を行わせる方法が提案された。これが GZ である。

GZ の最初の調査である Galaxy Zoo 1??では渦巻銀河か楕円銀河の区別や、渦巻銀河であった場合にどちら周りに渦巻いているかなどを投票形式で集計し、6 つのカテゴリへと分類が行われた。GZ1 の後継プロジェクトである Galaxy Zoo 2??では、GZ1 より詳細な形態分類を行うため 11 つの質問が用意されている。

GZ1 における形態ラベル付け方法

GZ1 では、ウェブサイト (www.galaxyzoo.org) を通じて有志の一般人の形態分類を投票形式で集計した。GZ1 における分類の投票ページを図 3.3 に示す。サイトを訪れた有志の一般人は、楕円銀河・時計回り渦巻銀河・反時計回り渦巻銀河・エッジオン (渦が地球から観測できない向きについている銀河)・星もしくは区別できなかった天体・マージャー (2 つの銀河が衝突し合体している, 合体銀河) の 6 つのうちいずれかに投票を行う。形態分類カタログは、対象となった銀河に対し、各形態分類の投票率を付与され作成される。

実験に用いる正解ラベルの作成方法

今回分類モデル学習に用いる正解ラベルとして、GZ1 から取得できる Table2 の分類フラグを用いた。Table2 の記載内容を示した図を図 3.4 に示す。Table2 は SDSS から取得した天体のうち、分光スペクトルデータが観測された天体に関して収録されている。Table2 には以下の情報が記載されている。

- SDSS における天体オブジェクト ID
- 天体の天球座標 (RA, Dec)
- 投票数



図 3.3: GZ1 における形態分類の投票ウェブページ
(??より引用)

- 各カテゴリの得票率
- バイアスが除去された投票率
- 分類フラグ (渦巻銀河・楕円銀河・不確かな天体)

分類フラグは、渦巻銀河・楕円銀河・不確かな天体の3つが存在する。それぞれの銀河種に対し、得票率が8割を超えた場合にフラグが立つ。今回はこの分類フラグを深層学習モデルの学習に用いる。

OBJID	RA	DEC	NVOTE	P_EL	P_CW	P_ACW	P_EDGE	P_DK	P_MG	P_CS	P_EL_DEBIASED	P_CS_DEBIASED	SPIRAL	ELLIPTICAL	UNCERTAIN
	hms	dms													
587727178986356823	00:00:00.41	-10:22:25.7	59	0.61	0.034	0.0	0.153	0.153	0.051	0.186	0.61	0.186	0	0	1

図 3.4: Table2

第4章 SDSS & Galaxy Zooを用いた分類モデル学習

この章では当論文の実験で用いられる SDSS と GZ にて、学習データとテストデータの解像度が揃っているという条件のもと、高精度分類が行える分類モデルを学習させることを目的とし、その結果 2 つの実験を行った。

第 4 章を行う動機は、当論文で掲げている将来展望の前提条件を達成することである。当論文の将来展望は、高空間分解能観測装置データを用いてモデル学習を行うことで、既存の低空間分解能データセットに対し更なる高精度形態分類を提供するというものである。この将来展望にまつわる実験の最も初段階の前提条件として、まずは学習データとテストデータの解像度が揃っている条件にて高精度分類が行えるかを検証する必要がある。そこで 2 つの実験を行った。

4.1 実験概要

第 4 章では SDSS から取得した銀河画像と GZ から取得した分類フラグを学習データとし、渦巻銀河・楕円銀河・不確かな天体のいずれかを予測する分類モデルを学習する。

学習データ

学習に用いる画像データは、SDSS から取得した 64 ピクセル四辺の銀河切り出し画像を使用し、正解ラベルは GZ から取得した分類フラグを用いる。実験には 15,000 天体を用いた。15,000 天体の赤方偏移別の個数を示したグラフを図 4.1 に示す。一般的に赤方偏移の値が小さいほど、地球から近い天体といえる。15,000 天体のうち、GZ の分類フラグにて渦巻銀河と分類されている天体は 4,058 天体、楕円銀河は 1,561 天体、不確かな天体は 9,310 天体であった。渦巻銀河・楕円銀河・不確かな天体の切り出し画像の例を図??から図??に示す。

モデルの学習および評価の際、モデルの学習データとテストデータの比率は 7:3 とした。

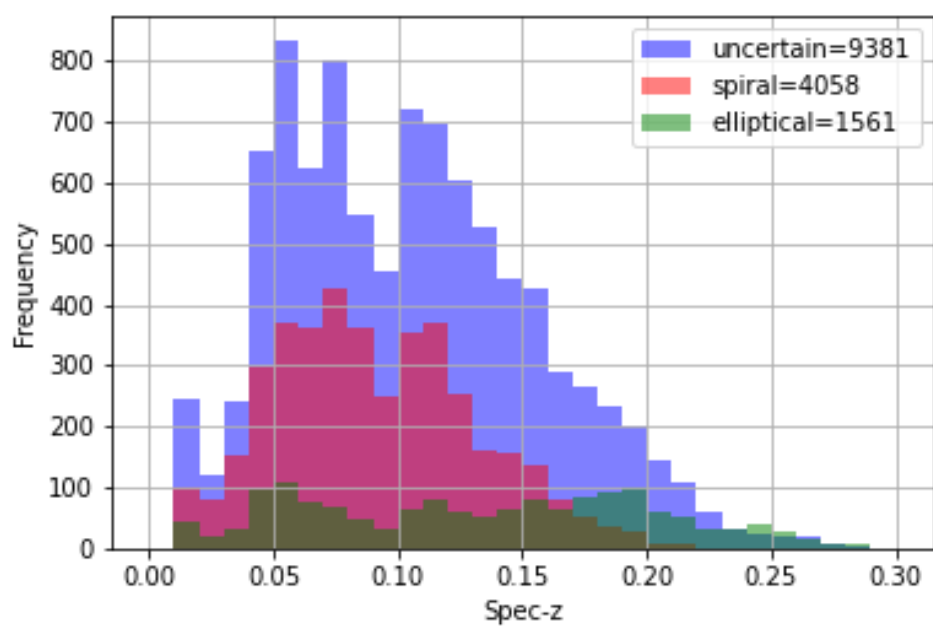
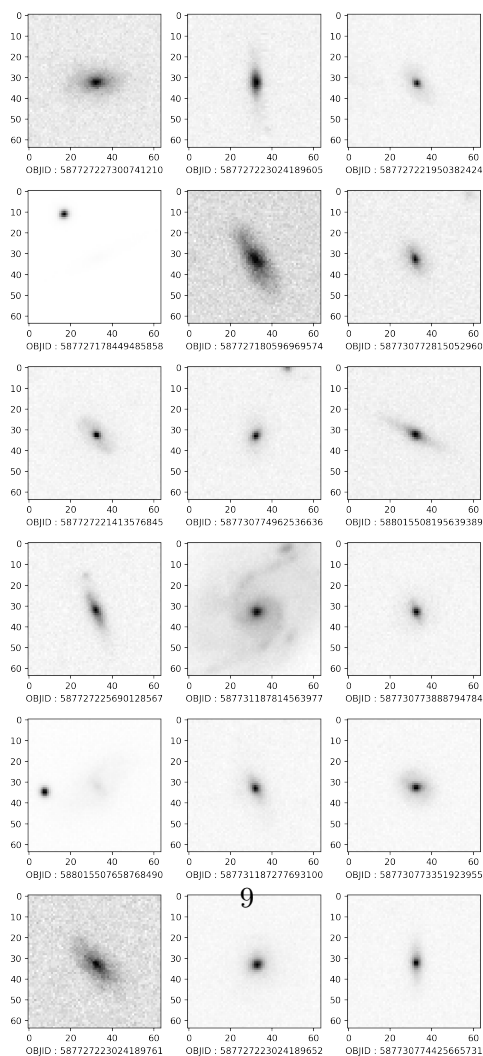
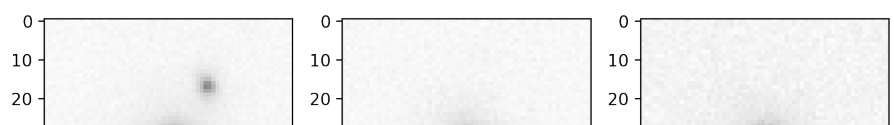
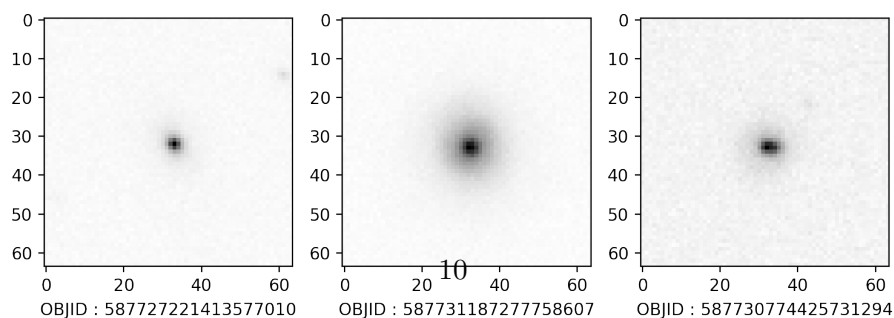
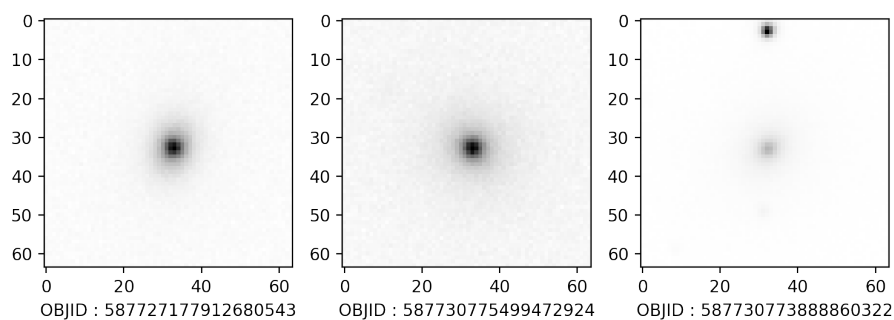
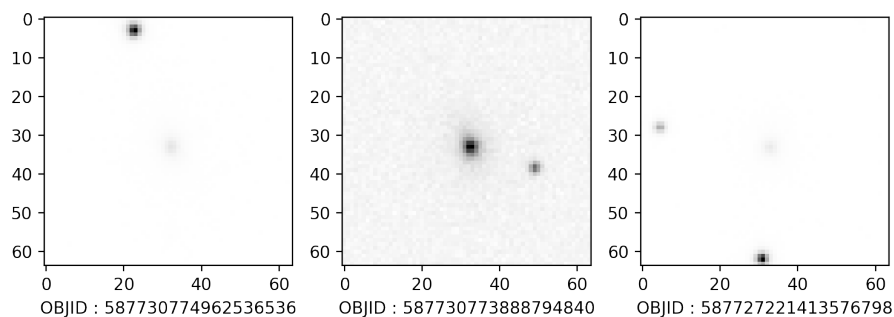
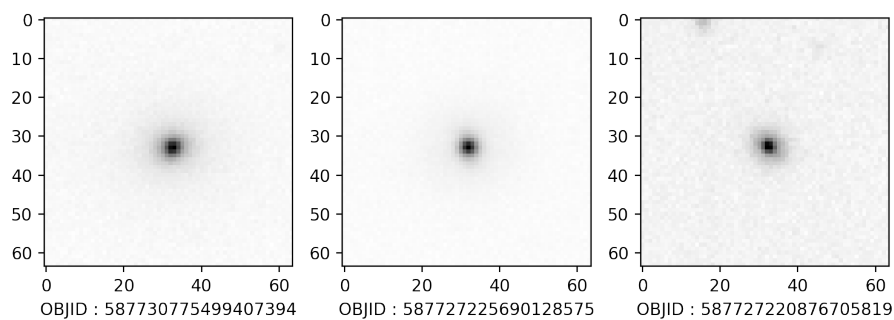
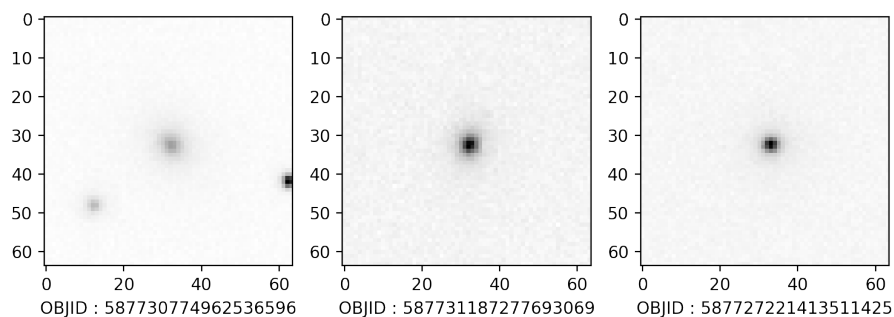


図 4.1: 実験に用いる 15,000 天体の赤方偏移別の個数グラフ





モデル構造

今回用いた深層学習モデルは、cheng et al.(2019)??にて用いられていた銀河形態分類モデルを参考にした。今回用いたモデルの構造を図 4.4 に示す。このモデルは畳み込み層を合計 3 つ有しており、それぞれのカーネルサイズは 3×3 , 3×3 , 2×2 である。それぞれの畳み込み層の後には、 2×2 の max-pooling 層が存在する。全畳み込み層の後に全結合層が 2 層配置されており、それぞれ 1024 個のノードを有している。

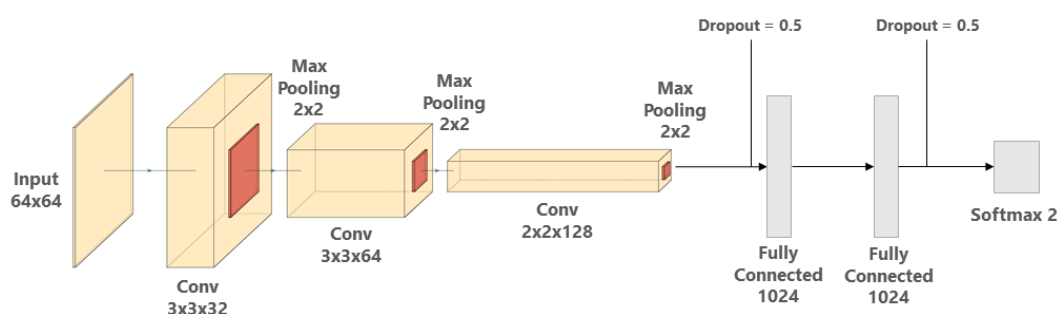


図 4.4: 用いた分類モデルの構造図

モデルの評価方法

分類モデルの評価指標として、主に accuracy (正解率) と true positive rate (真陽性率), そして一部実験において True Skill Statistics (以下, TSS) を用いた。accuracy は全テストデータの中で正しく分類できたデータがどれだけあるかというものであり、モデルの正確性を表す指標である。true positive rate は本来陽性だと分類すべき全データのうち、どれほど正しく分類できたかを表す指標である。TSS も accuracy と同じく正確性を表す指標であるが、テストデータ内のデータインバランス性に対しロバストな性質がある。

以下に 2 値分類問題を例とした場合の混同行列, accuracy, true positive rate および TSS の導出方法を示す。ここで混同行列とは、分類問題においてモデルが予測した値および真の値を行列形式で表したものであり、分類モデルの性能を可視化・評価するのによく用いられる指標である。モデルの予測値と真の値が交差する対角成分における数が多いほど、モデルが正確な予測を行っているといえる。

		モデルの予測値	
		陽性 (Positive)	陰性 (Negative)
実際の値	陽性 (Positive)	TP (True Positive)	FN (False Negative)
	陰性 (Negative)	FP (False Positive)	TN (True Negative)

図 4.5: 2 値分類問題における混同行列

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.1)$$

$$truepositiverate = \frac{TP}{TP + FN} \quad (4.2)$$

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{TN + FP} \quad (4.3)$$

また、深層学習モデルには実行毎に学習のブレが存在するため、モデルの評価の際には学習・テストを 30 回実行し、評価指標の平均値、標準偏差および標準誤差を導出する。

4.2 不確かラベルが与える分類精度への影響

GZ が与える形態分類フラグのうち「不確かな天体」というフラグが立っている天体は、渦巻銀河・楕円銀河のどちらの得票率も 8 割を超えなかった天体である。これらの天体は人間の目による形態分類が比較的難しい天体、つまり天体画像から読み取れる形態的特徴があいまいな天体であると考えられる。

この節では特徴があいまいだと考えられる不確かな天体群が、形態分類モデルの分類精度に与える影響を調べる。具体的には以下の 2 条件でモデルの学習・テストを行い、テストデータに対

する予測の accuracy や混同行列の比較をする．そして不確かな天体の特性や学習に用いた際に分類精度に与える影響を調べる．

- 不確かを含めた，渦巻・楕円・不確かの 3 値分類
- 不確かを除いた，渦巻・楕円の 2 値分類

4.2.1 実験条件

モデルの学習およびテストを行う天体数は 1,000 天体とした．使用する天体は，GZ による形態分類が行われた 15,000 天体 (図 4.1 参照) からランダムに取得を行う．取得を行う際，赤方偏移 z について， $0 < z < 0.2$ という条件を設定し，あてはまらない天体は除外を行う．これは，地球との距離が遠すぎる天体については上手く特徴抽出が行えず学習の妨げになる可能性があるからである．

使用天体を 1,000 天体取得したあと，学習データとテストデータの比率が 7:3 となるように切り分けを行った．

モデルの評価を行う際，モデルの学習およびテストを 30 回行い，accuracy の平均値，標準偏差および標準誤差を導出した．なお，30 回の学習およびテストの際，学習実行毎に取得される天体は毎回シャッフルされる．

4.2.2 実験結果

学習結果

不確かを含めた渦巻・楕円・不確かの 3 値分類モデル，および不確かを除いた渦巻・楕円の 2 値分類モデルの学習結果を図 4.6 から図 4.11 に示す．

- 一番上の図は，横軸 epoch 数・縦軸 loss (損失関数) となっている．
- 中段の図は，横軸 epoch 数・縦軸 accuracy (正解率) となっている．
 - これら 2 つの図は，青色のラインが学習データに対するスコア，黄色のラインがテストデータに対するスコアとなっている．
 - loss のグラフについて，学習データに対する損失関数は学習が進むにつれ順当に下がっていくものの，ある一定の epoch からテストデータに対する損失関数が上昇していく現象が見受けられる．この現象は過学習と呼ばれている．この現象が起こると，学習データに対し

分類モデルが過剰に適合した結果、学習データに対する予測精度に
比ベテストデータに対する予測精度が低下していく。

- 一番下の図はテストデータに対する予測結果から生成した混同行列であり、行成分がモデルによる予測ラベル、列成分がテストデータの真のラベルである。

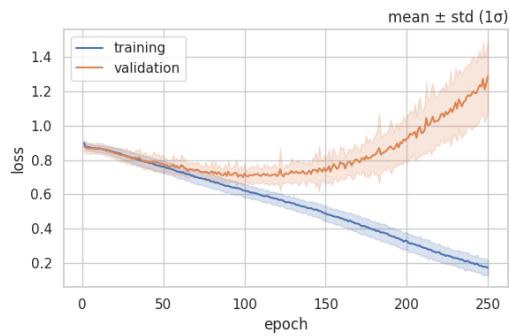


図 4.6: 3 値分類 : loss 関数の学習遷移

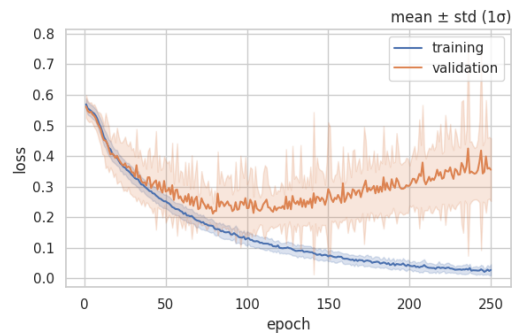


図 4.7: 2 値分類 : loss 関数の学習遷移

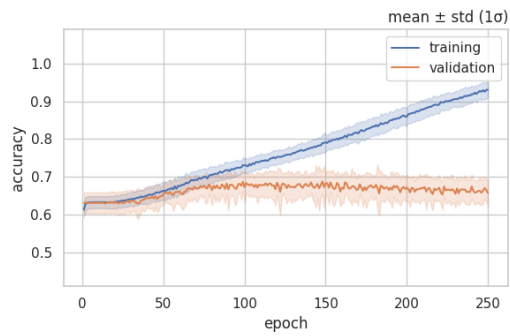


図 4.8: 3 値分類 : accuracy の学習遷移

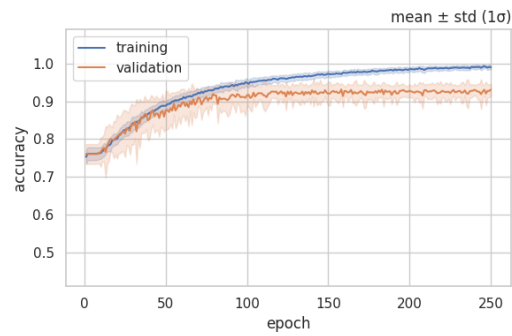


図 4.9: 2 値分類 : accuracy の学習遷移

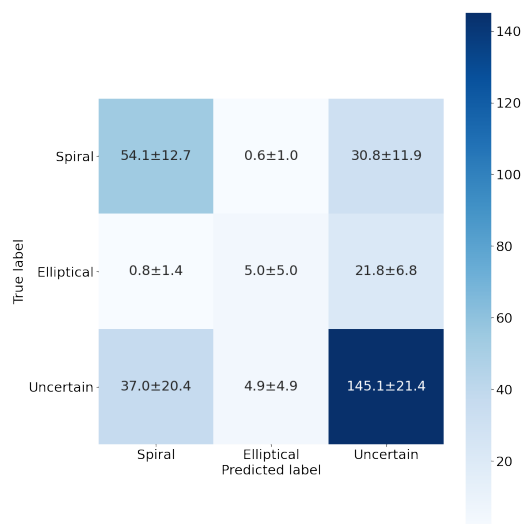


図 4.10: 3 値分類：混同行列
(平均 ± 標準偏差 (1σ))

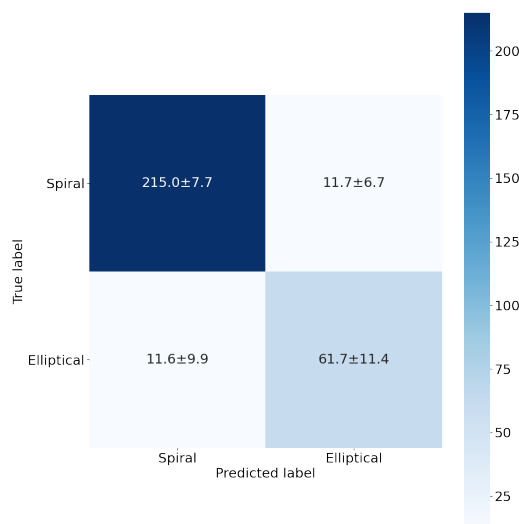


図 4.11: 2 値分類：混同行列
(平均 ± 標準偏差 (1σ))

結果

モデル学習の進行 図 4.6・図 4.8 において，テストデータに対するスコアが 100epoch 付近まで loss 関数は降下，accuracy は上昇し，それ以降の epoch では loss 関数が上昇，accuracy が降下していることから，3 値分類モデルでは 100epoch まで学習が良い方向に進んでいるもののそれ以降の epoch からはモデルが学習データに対し過学習し，テストデータに対する汎化性能を失っていることが分かる．

また図 4.7・図 4.9 において，テストデータに対するスコアが 100epoch 付近まで loss 関数は降下，accuracy は上昇するものの，それ以降の epoch では loss 関数は上昇するが accuracy は横ばいになっている．このことから，2 値分類モデルにおいては 100epoch 目以降学習データに対し過適合はしているものの，テストデータに対する汎化性能は失っていないことが読み取れる．

テストデータに対する accuracy の評価 モデルの形態分類性能を評価するため，両モデル間にてテストデータに対する accuracy の比較を行う．性能評価においては，モデルが学習しきった際の分類性能を用いる．すなわち，accuracy を採る epoch 数は，テストデータに対する accuracy が最大となる epoch 数とした．

不確かを含めた渦巻・楕円・不確かの 3 値分類モデル，および不確かを除いた渦巻・楕円の 2 値分類モデルにおける，テストデータへの予測精度 (accuracy) を表 4.1 に示す．3 値分類・2 値分類ともに，accuracy を採る epoch 数を表内に示している．表 4.1 において，1 行目には今回の 30 回

実行における accuracy の平均値と標準偏差, 2 行目には accuracy の平均値と今回の 30 回実行から推定した標準誤差をそれぞれ 2σ 分記載した.

表 4.1 より, 標準誤差付き平均値において 3 値分類モデルと 2 値分類モデルとの間に有意な差があり, なおかつ 2 値分類モデルの方が accuracy が高いことが分かる. また標準偏差付き平均値においても 3 値分類モデルと 2 値分類モデルとの間に有意な差があることから, 両モデルとも学習のブレがさほど大きくないことが分かる.

表 4.1: 不確か天体を含めた 3 値分類, および不確か天体を除いた 2 値分類モデルのテストデータに対する予測の accuracy

	mean \pm std (2σ)	mean \pm ste (2σ)
3 値分類 (渦巻・楕円・不確か)	0.688 \pm 0.062(148th epoch)	0.688 \pm 0.011(148th epoch)
2 値分類 (渦巻・楕円)	0.931 \pm 0.034(158th epoch)	0.931 \pm 0.006(158th epoch)

混同行列の導出方法, および両モデルにおける比較 混同行列を導出するため, 先ほど求めた「モデルが学習しきった地点と思われる, テストデータに対する accuracy が最大となる epoch 数」まで, 再度 30 回学習およびテストを 3 値分類モデル・2 値分類モデル両方にて行った. 導出された混同行列をそれぞれ図 4.10, 図 4.11 に示す. 図にはそれぞれ個数の平均値と標準偏差が 1σ で示されている.

図 4.10 より, 3 値分類モデルにおいて陽性を不確かな天体とした場合の true positive rate は約 75%だが, 渦巻銀河群においては約 60%, 楕円銀河群においては約 20%である. このことから, 学習した 3 値分類モデルは不確かな天体群に比べ渦巻・楕円銀河を正しく分類する性能が劣っており, 特に楕円銀河を正しく分類性能することができないことが読み取れる. 図 4.11 より, 2 値分類モデルにおいて陽性を渦巻銀河とした場合の true positive rate は約 90%, 楕円銀河とした場合は約 80%であることが分かる. このことから, 学習した 2 値分類モデルにおいて渦巻・楕円銀河を正しく分類する性能は 3 値分類モデルよりも優れていることが分かる.

また, 図 4.10, 図 4.11 より, 3 値分類モデルに用いたテストデータ内には不確かな天体群が約 2/3 存在し, また 2 値分類モデルに用いたテストデータ内には渦巻銀河と楕円銀河の比率がおおよそ 3:1 となっていることがわかる. これは今回用いる天体の母集団である 15,000 天体 (図 4.1 参照) の特性と一致している.

4.3 データインバランスが与える分類精度への影響

4.2 節において、用いる銀河種を渦巻銀河・楕円銀河の 2 値とすることで、分類モデルの精度向上が行えることがわかった。一方、図 4.10, 4.11 より、テストデータにおいて銀河種のデータインバランスが起きていることがわかった。このインバランス性は今回のモデル学習・テストに用いている銀河の母集団 (図 4.1 参照) の性質と一致する。

この節では、学習データ・テストデータ内の銀河種におけるデータインバランス性が、形態分類モデルの分類精度に与える影響を調べる。具体的には以下の 2 条件でモデルの学習・テストを行い、テストデータに対する予測の accuracy, TSS や混同行列の比較をする。そしてデータインバランス性が分類精度に与える影響を調べる。

- 渦巻銀河・楕円銀河の比率が異なるインバランسدデータで学習・テストを行った 2 値分類モデル
- 渦巻銀河・楕円銀河の比率が等しいバランسدデータで学習・テストを行った 2 値分類モデル

4.3.1 実験条件

モデルの学習およびテストを行う天体の選定方法として、インバランسدデータにおいては GZ による形態分類が行われた 15,000 天体 (図 4.1 参照) のうち不確かな天体を除いた渦巻銀河・楕円銀河からランダムに 1,000 天体の取得を行った。またバランسدデータにおいては渦巻銀河と楕円銀河の比率を等しくするために、15,000 天体の母集団のうち不確かな天体を除いた渦巻銀河・楕円銀河から、渦巻銀河を 500 天体、楕円銀河を 500 天体取得した。なお取得を行う際、赤方偏移 z について、 $0 < z < 0.2$ という条件を設定し、あてはまらない天体は除外を行う。これは、地球との距離が遠すぎる天体については上手く特徴抽出が行えず学習の妨げになる可能性があるからである。

使用天体を 1,000 天体取得したあと、学習データとテストデータの比率が 7:3 となるように切り分けを行った。このうちバランسدデータでは、学習データ・テストデータのどちらも渦巻銀河と楕円銀河の比率が等しくなるように 1,000 天体からの切り分けを行った。

モデルの評価を行う際、モデルの学習およびテストを 30 回行い、accuracy の平均値、標準偏差および標準誤差を導出した。なお、30 回の学習およびテストの際、学習実行毎に使用される天体は毎回シャッフルされる。また 4.3 節の実験においては、モデルの評価指標として TSS の導出も行った。これはインバランسدデータでのモデル学習・テストにおいて、30 回実行における実行

毎に取得する天体の渦巻銀河・楕円銀河の比率が毎回異なることから、accuracy だけでは銀河種のデータインバランス性のみを正しく評価できないためである。そこでデータセット内の分類ラベルのインバランス性にロバストな性質をもつ TSS も評価指標に加えている。

4.3.2 実験結果

学習結果

渦巻銀河・楕円銀河の比率が異なるインバランسدータで学習を行った 2 値分類モデル、および渦巻銀河・楕円銀河の比率が等しいバランسدータで学習を行った 2 値分類モデルの学習結果を図 4.12 から図 4.17 に示す。

- 一番上の図は、横軸 epoch 数・縦軸 loss（損失関数）となっている。
- 中段の図は、横軸 epoch 数・縦軸 accuracy（正解率）となっている。
 - － これら 2 つの図は、青色のラインが学習データに対するスコア、黄色のラインがテストデータに対するスコアとなっている。
 - － loss のグラフについて、学習データに対する損失関数は学習が進むにつれ順当に下がっていくものの、ある一定の epoch からテストデータに対する損失関数が上昇していく現象が見受けられる。この現象は**過学習**と呼ばれている。この現象が起こると、学習データに対し分類モデルが過剰に適合した結果、学習データに対する予測精度に比べテストデータに対する予測精度が低下していく。
- 一番下の図はテストデータに対する予測結果から生成した混同行列であり、行成分がモデルによる予測ラベル、列成分がテストデータの真のラベルである。

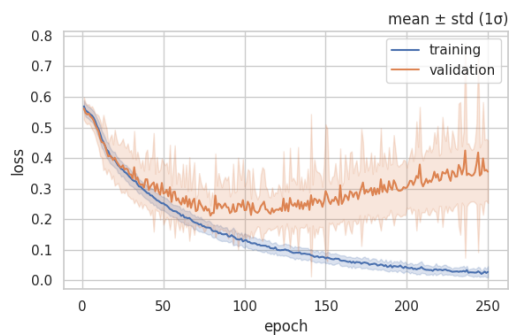


図 4.12: インバランسدデータ : loss 関数の学習遷移

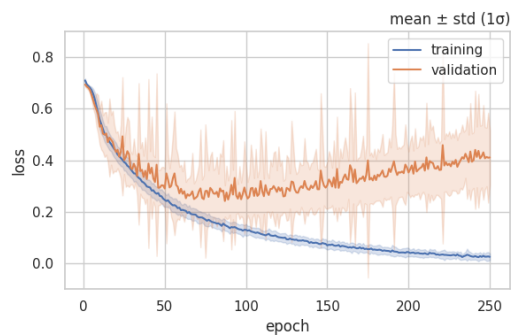


図 4.13: バランسدデータ : loss 関数の学習遷移

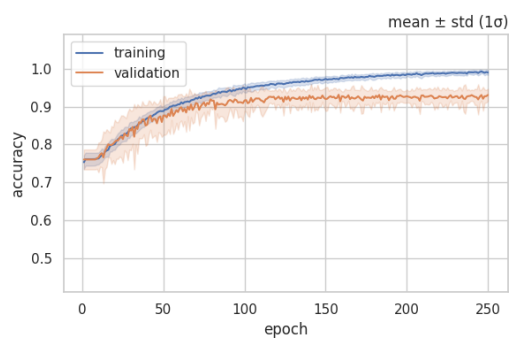


図 4.14: インバランسدデータ : accuracy の学習遷移

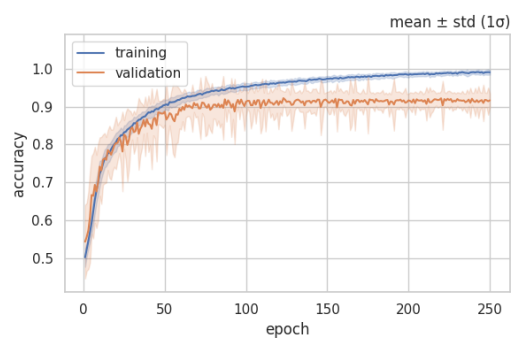


図 4.15: バランسدデータ : accuracy の学習遷移

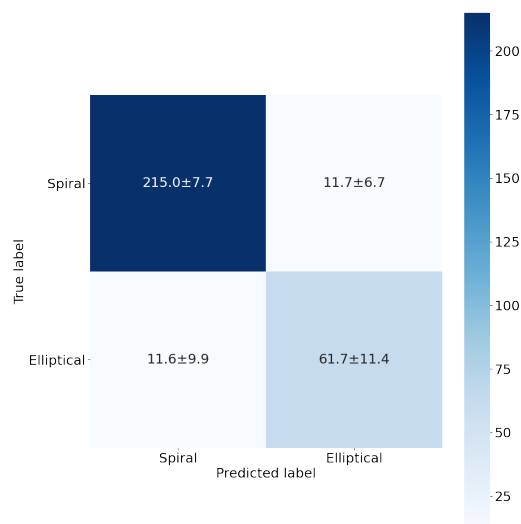


図 4.16: インバランسدデータ：混同行列
(平均 ± 標準偏差 (1σ))

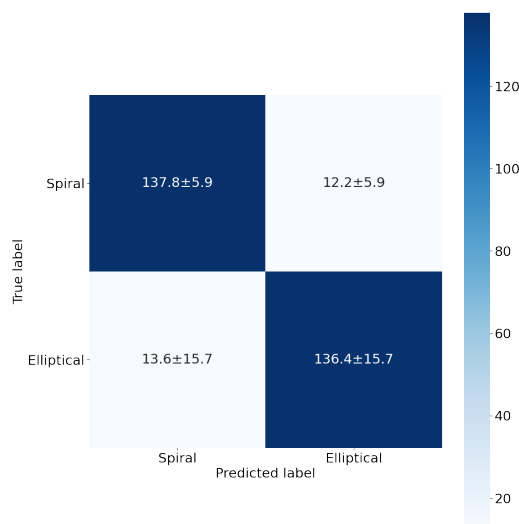


図 4.17: バランسدデータ：混同行列
(平均 ± 標準偏差 (1σ))

結果

モデル学習の進行 図 4.12・図 4.14 において、テストデータに対するスコアが 100epoch 付近まで loss 関数は降下、accuracy は上昇するものの、それ以降の epoch では loss 関数は上昇するが accuracy は横ばいになっている。このことから、インバランسدデータで学習を行ったモデルは 100epoch 目以降学習データに対し過適合はしているものの、テストデータに対する汎化性能は失っていないことが読み取れる。

また図 4.13・図 4.15 において、テストデータに対するスコアが 75epoch 付近まで loss 関数は降下、accuracy は上昇するものの、それ以降の epoch では loss 関数は上昇するが accuracy は横ばいになっている。このことから、バランسدデータで学習を行ったモデルは 75epoch 目以降学習データに対し過適合はしているものの、テストデータに対する汎化性能は失っていないことが読み取れる。

テストデータに対する accuracy の評価 モデルの形態分類性能を評価するため、両モデル間にてテストデータに対する予測の accuracy の比較を行う。性能評価においては、モデルが学習しきった際の分類性能を用いる。すなわち、accuracy を採る epoch 数は、テストデータに対する accuracy が最大となる epoch 数とした。

インバランسدデータで学習を行ったモデル、およびバランسدデータで学習を行ったモデルにおける、テストデータへの予測精度 (accuracy) を表 4.2 に示す。両モデルにて、accuracy を採

る epoch 数を表内に示している。表 4.2 において、1 行目には今回の 30 回実行における accuracy の平均値と標準偏差、2 行目には accuracy の平均値と今回の 30 回実行から推定した標準誤差をそれぞれ 2σ 分記載した。

表 4.2 より、テストデータに対する予測の accuracy の標準誤差付き平均値においてインバランスデータで学習を行ったモデルとバランسدデータで学習を行ったモデルとの間に有意な差はなかった。

TSS の導出方法、および両分類モデルの分類性能評価 データセット内の分類ラベルのインバランス性に左右されづらい評価指標である TSS にて、再度両モデルの評価・性能比較を行う。

TSS を導出するため、先ほど求めた「モデルが学習しきった地点と思われる、テストデータに対する accuracy が最大となる epoch 数」まで、再度 30 回学習およびテストを両モデルにて行った。今回行ったこの再度 30 回学習およびテストをした結果は、後ほど混同行列の導出にも用いる。

インバランسدデータで学習を行ったモデル、およびバランسدデータで学習を行ったモデルにおける、テストデータへの予測精度 (TSS) を表 4.3 に示す。1 行目には今回の 30 回実行における TSS の平均値と標準偏差、2 行目には TSS の平均値と今回の 30 回実行から推定した標準誤差をそれぞれ 2σ 分記載した。表 4.3 より、標準誤差付き平均値を見たとき、TSS が両モデルとも 0.8 付近と高いスコアを記録しており、また両モデルにおける有意差はなかった。このことから、両モデルとも渦巻銀河と楕円銀河をそれぞれ正しく分類する能力を持っており、またモデル学習において渦巻銀河・楕円銀河どちらの特徴もある程度正しく学習できていることが分かる。

混同行列によるモデルの評価 TSS 導出に用いた「モデルが学習しきった地点と思われる、テストデータに対する accuracy が最大となる epoch 数まで、再度モデルの 30 回学習及びテスト」の結果を用い、両モデルの混同行列を導出した。導出された混同行列をそれぞれ図 4.16、図 4.17 に示す。図にはそれぞれ個数の平均値と標準偏差が 1σ で示されている。

図 4.16 より、インバランسدデータを用いて学習を行ったモデルにおいて陽性を渦巻銀河とした場合の true positive rate は 0.948、楕円銀河とした場合は 0.842 であることが分かる。また図 4.17 より、バランسدデータを用いて学習を行ったモデルにおいて陽性を渦巻銀河とした場合の true positive rate は 0.919、楕円銀河とした場合は 0.909 であることが分かる。

4.4 4 章全体の議論・結論

4.2 節の実験にて、不確かな天体を入れた 3 値分類問題とした方がテストデータに対する予測精度が落ちてしまう現象が確認された。この現象の理由としては、不確かな天体が GZ にて形態ラ

表 4.2: インバランسدデータで学習を行ったモデル, およびバランسدデータで学習を行ったモデルのテストデータに対する予測の accuracy

	mean \pm std (2σ)	mean \pm ste (2σ)
インバランسدデータでのモデル	0.931 \pm 0.034(158th epoch)	0.931 \pm 0.006(158th epoch)
バランسدデータでのモデル	0.922 \pm 0.036(123th epoch)	0.922 \pm 0.007(123th epoch)

表 4.3: インバランسدデータで学習を行ったモデル, およびバランسدデータで学習を行ったモデルのテストデータに対する予測の TSS

	mean \pm std (2σ)	mean \pm ste (2σ)
インバランسدデータでのモデル	0.791 \pm 0.250	0.791 \pm 0.046
バランسدデータでのモデル	0.828 \pm 0.184	0.828 \pm 0.033

ベル付けを行った人間にとっては分類できず, 深層学習分類モデルにとっては分類できるほどの特徴を有していることが推測される. 分類モデルには渦巻銀河・楕円銀河のどちらかに分類できる天体に「不確かな天体」というラベル付けがされているため, モデルが渦巻銀河・楕円銀河・不確かな天体を3つに分類できるような学習がされないのではないかと考えられる. もし不確かな天体と分類されている天体に正確なラベル付けが行われれば, 深層学習分類モデルの分類精度は向上するものと思われる.

4.3 節の実験にて, 学習・テストに用いるデータ内のインバランス改善を行っても精度が改善しなかった理由について, これはインバランسدデータ内にてよりデータ数の少ない楕円銀河においても, テストデータ内の楕円銀河を分類する能力獲得に十分なデータ数および特徴の多様性が存在していたことが考えられる. もし学習データ内にて桁数が異なるほどの強いデータインバランスが起きていた場合, データインバランス改善を行えば分類精度が改善される可能性がある.

4 章の目標は「学習データとテストデータの解像度が等しい」という条件にて, 高精度分類が行えるモデルを学習させることであった. この目標を達成するため, 2つの実験を行った. GZにて不確かと分類されている天体を除外し, 渦巻銀河・楕円銀河の2値分類問題とすることで, 分類モデルの分類精度は大きく向上した. また, 学習・テストに用いるデータ内のデータインバランスを改善しても分類精度に有意差が認められるほどの向上は見受けられなかった. 当論文における今後の実験では, 渦巻銀河を見分けることに特化したモデルでなく, 渦巻銀河・楕円銀河の両者を分類することを優先するため, 渦巻銀河・楕円銀河の両者の真陽性率向上が見込めるバランسدデータにてモデル学習を行っていく.

第5章 空間解像度差のあるデータセットを用いた 分類モデル学習

当論文の将来展望は、高空間分解能観測装置データを用いてモデル学習を行うことで、既存の低空間分解能データセットに対し更なる高精度形態分類を提供するというものである。この将来展望の前提条件である、学習データとテストデータとの間の解像度が揃っているデータセットにて高精度の銀河形態分類モデルを学習させることを、第4章にて達成することができた。

そこで、この章では将来展望で行われる予定の「異なる観測装置データの組み合わせ」を念頭に置き、学習データとテストデータとの間に解像度差があるデータセットにて形態分類モデルの学習が行えるかを検証する。

具体的には、以下の事項を検証するため、2つの実験を行った。

- 学習データとテストデータとの間に解像度差があるデータセットにて形態分類モデルを学習し、
 - － テストデータへの予測が成立するか
 - － 予測が成立する場合、解像度差がいくつまでならば使用可能か

5.1 実験概要

第5章ではSDSSから取得した銀河切り出し画像とGZから取得した分類フラグを学習データとしてモデルを学習させ、銀河切り出し画像を縮小した低解像度データセットに対し予測を行う。なお、学習・テストに用いる天体データの銀河種は渦巻銀河・楕円銀河の2つとする。

学習データ

学習に用いる画像データは、SDSSから取得した64ピクセル四辺の銀河切り出し画像を使用し、正解ラベルはGZから取得した分類フラグを用いる。実験には15,000天体を用いた。15,000天体の赤方偏移別の個数を示したグラフを図5.1に示す。一般的に赤方偏移の値が小さいほど、地球か

ら近い天体といえる。15,000 天体のうち、GZ の分類フラグにて渦巻銀河と分類されている天体は 4,058 天体、楕円銀河は 1,561 天体であった。

モデルの学習および評価の際、モデルの学習データとテストデータの比率は 7:3 とした。

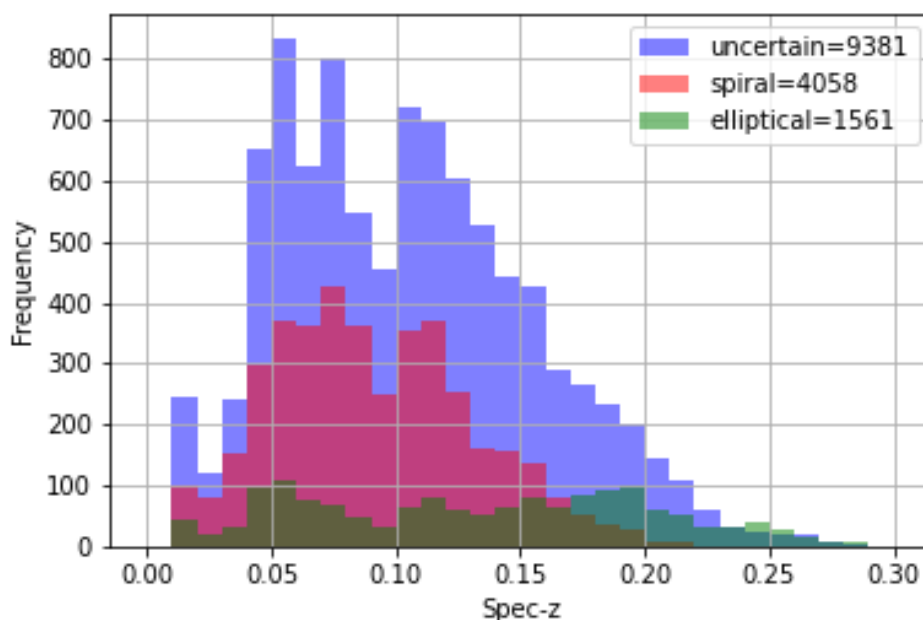


図 5.1: 実験に用いる 15,000 天体の赤方偏移別の個数グラフ

縮小画像データセット (テストデータ) の作成方法

この章ではテストデータとして、SDSS から取得した銀河切り出し画像を縮小したデータセットを使用する。SDSS から取得した 15,000 天体 (図 5.1 参照) を平均画素法により縮小処理を施した。画像の縮小には OpenCV?? に実装されている `resize` 関数の "INTER_AREA" メソッドを使用した。銀河切り出し画像を 1/2 倍から 1/8 倍に縮小し、計 7 つの縮小画像データセットを作成した。縮小画像の例として、1/4 倍に縮小した画像群を図 5.2 に示す。

縮小画像データセットの内挿方法

モデルに画像データを入力する際、用いる画像サイズは揃えて入力する必要がある。第 5 章では学習データとテストデータとの間で画像サイズが異なる。そのため、今回の実験ではテストデータである縮小画像データセットを拡大してモデルに入力する。

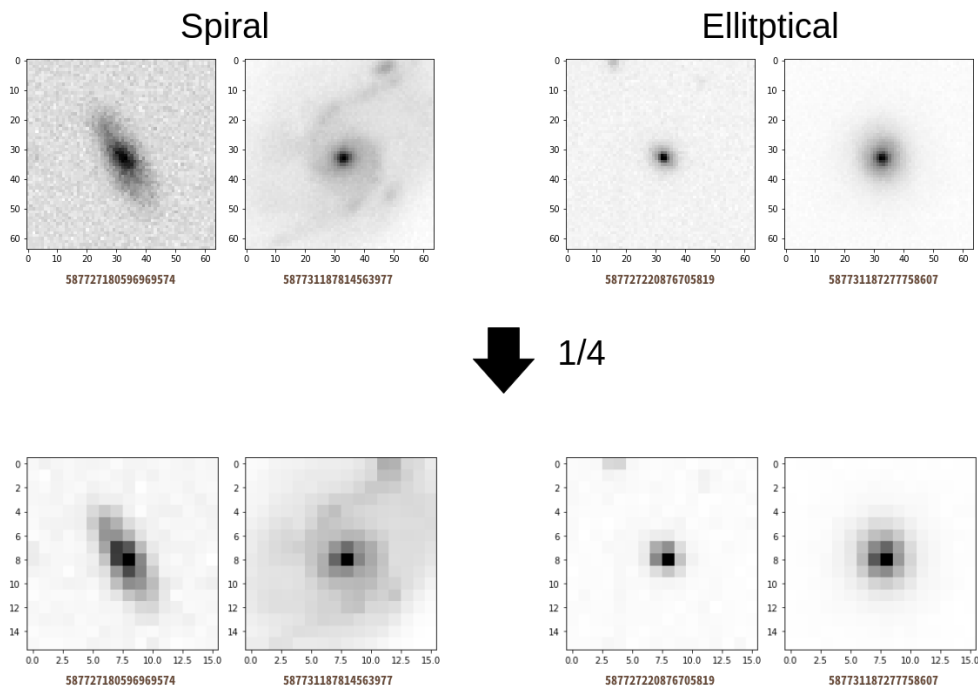


図 5.2: 銀河切り出し画像の縮小例 (1/4 倍)

縮小画像の拡大には、OpenCV の `resize` 関数を用いた。この章での実験において、画像拡大の際の補間方法として最近傍補間、バイリニア補間、バイキュービック補間を使用した。画像拡大の例として、1/4 倍に縮小された銀河切り出し画像の、各補間方法による画像拡大の例を図 5.3 に示す。なお図 5.3 にて用いられている縮小画像は、図 5.2 のものと同様である。

モデル構造

今回用いた深層学習モデルは、cheng et al.(2019)??にて用いられていた銀河形態分類モデルを参考にした。今回用いたモデルの構造を図 5.4 に示す。このモデルは畳み込み層を合計 3 つ有しており、それぞれのカーネルサイズは 3×3 , 3×3 , 2×2 である。それぞれの畳み込み層の後には、 2×2 の max-pooling 層が存在する。全畳み込み層の後に全結合層が 2 層配置されており、それぞれ 1024 個のノードを有している。

モデルの評価方法

分類モデルの評価指標として、accuracy (正解率) を用いた。accuracy は全テストデータの中で正しく分類できたデータがどれだけあるかというものであり、モデルの正確性を表す指標である。

また、深層学習モデルには実行毎に学習のブレが存在するため、モデルの評価の際には学習・テストを 30 回実行し、評価指標の平均値、標準偏差および標準誤差を導出する。

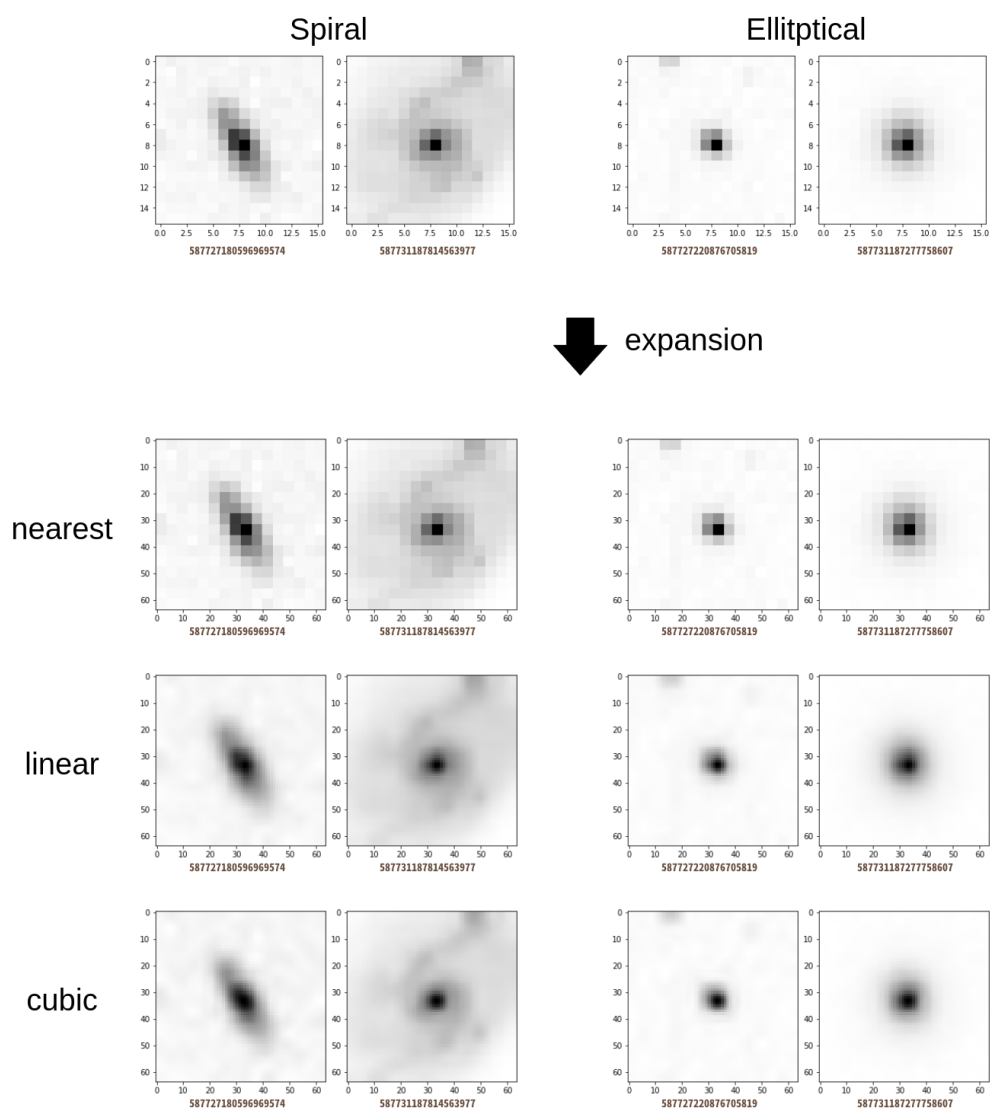


図 5.3: 1/4 倍縮小画像の画像拡大例//(最近傍補間・バイリニア補間・バイキュービック補間)

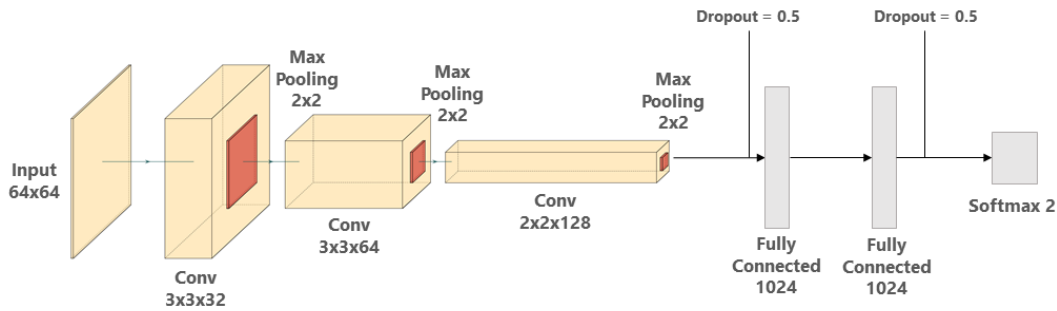


図 5.4: 用いた分類モデルの構造図

5.2 分類精度のデータセット内解像度差依存性

この節では、学習データとテストデータとの間に解像度差がある状況において、(1) テストデータへの予測が成立するか(2) 予測が成立する場合、解像度差がいくつまでならば使用可能かの検証を行う。このうち (2) において「使用可能」の定義として、学習データとテストデータの解像度が揃っている状況での予測結果である”Original”と同程度の高精度な分類、すなわち accuracy が 0.9 付近を記録する結果であることとする。

5.2.1 実験条件

使用する天体は、GZ による形態分類が行われた 15,000 天体 (図 4.1 参照) を母集団とした。銀河種類ごとのデータインバランスが起らないようにするため、その母集団の中から渦巻銀河と楕円銀河をそれぞれ 300 天体選定した。選定を行う際、赤方偏移 z について、 $0 < z < 0.2$ という条件を設定し、あてはまらない天体は除外を行う。これは、地球との距離が遠すぎる天体については上手く特徴抽出が行えず学習の妨げになる可能性があるからである。

使用天体を渦巻銀河 300 天体、楕円銀河 300 天体の合計 600 天体を選定したあと、学習データとテストデータの比率が 7:3 となるように天体群を分けた。このうち、学習データには SDSS の切り出し画像、テストデータには SDSS の切り出し画像を縮小した画像をそれぞれ用いた。また、学習データとテストデータ内において、銀河種のデータインバランスが起らないようにするため、渦巻銀河と楕円銀河の比率が等しくなるようにした。

モデルの評価を行う際、モデルの学習およびテストを 30 回行い、accuracy の平均値、標準偏差および標準誤差を導出した。なお、30 回の学習およびテストの際、学習実行毎に使用される天体は毎回シャッフルされる。

モデルの学習期間は 900epoch とした。これは、すべての縮小データの縮小倍率にて、モデルが学習しきる期間、すなわちテストデータに対する accuracy が上がりきる epoch 数が 900epoch で

あったからである.

5.2.2 実験結果

縮小データに対する予測結果

銀河切り出し画像を $1/2$ 倍から $1/8$ 倍に縮小した縮小データに対する, 形態分類モデルの予測結果の図を図 5.5 から図 5.8 に示す.

- 横軸が, テストデータである縮小データの縮小倍率であり, 縦軸がテストデータに対する予測の accuracy となっている.
 - テストデータに対する予測の accuracy について, 30 回実行における平均値と標準偏差 (2std) をエラーバー形式で示している.
 - なお, 一番左に位置している Original となっている箇所は, テストデータに縮小データでなく, 学習データと同じ大きさの銀河切り出し画像を使用してテストを行った結果である. すなわち, Original が学習データとテストデータの解像度が揃っている状況での予測結果に対し, $1/2$ から $1/8$ が学習データとテストデータとの間に解像度差がある状況での予測結果を示している.
- accuracy を採る epoch 数は, モデルが十分に学習しきる epoch 数とした. すなわち, テストデータに対する accuracy が最大値となる epoch 数とした.
- 図 5.8 には, 内挿方法間での accuracy の比較を図に示した.

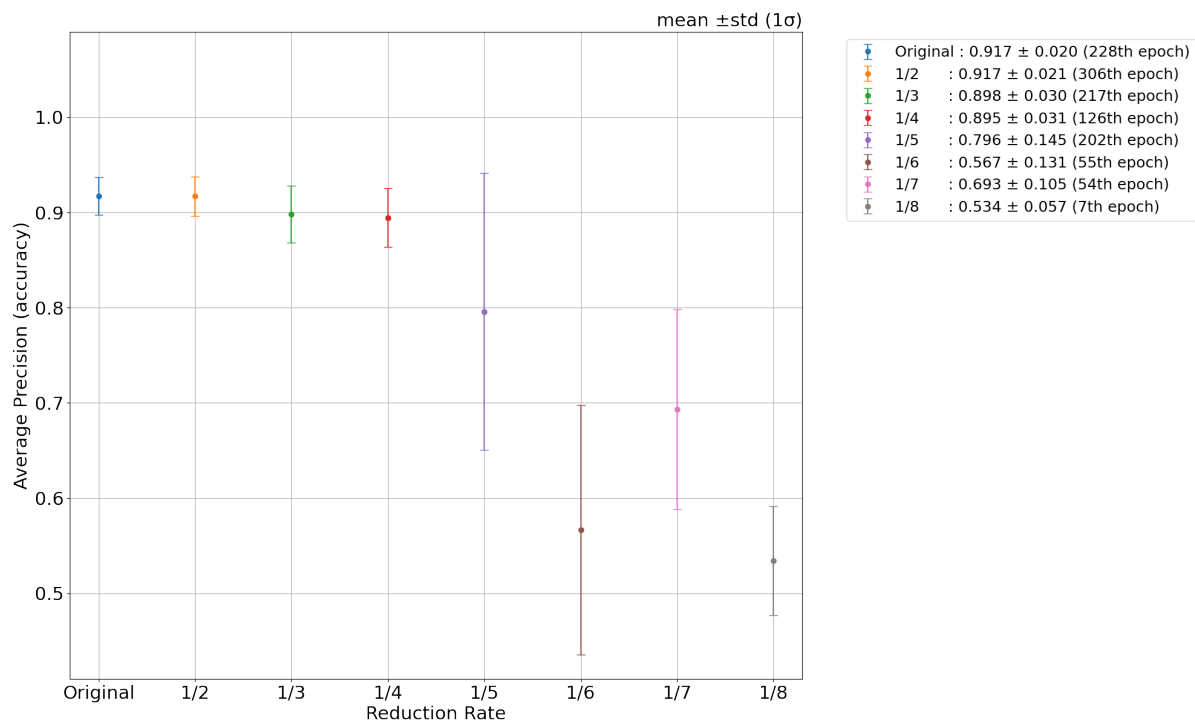


図 5.5: 最近傍補間によって内挿された, 縮小データに対するモデルの予測結果 (1std)

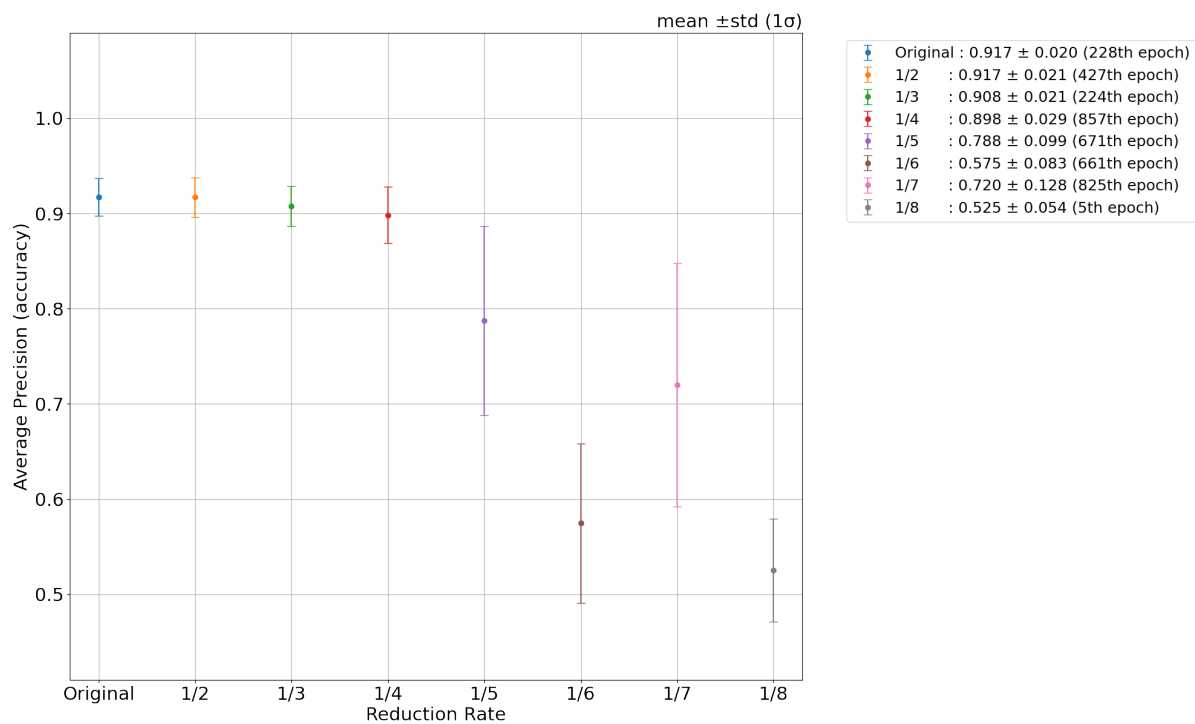


図 5.6: バイリニア補間によって内挿された, 縮小データに対するモデルの予測結果 (1std)

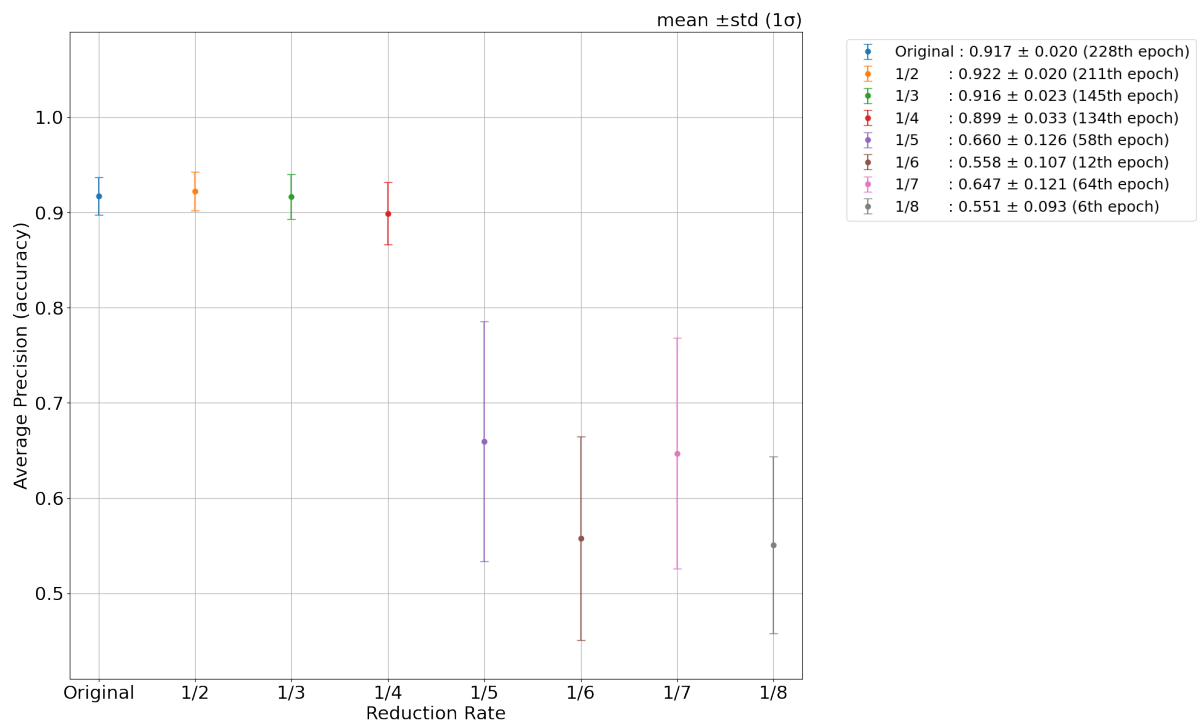


図 5.7: バイキュービック補間によって内挿された、縮小データに対するモデルの予測結果 (1std)

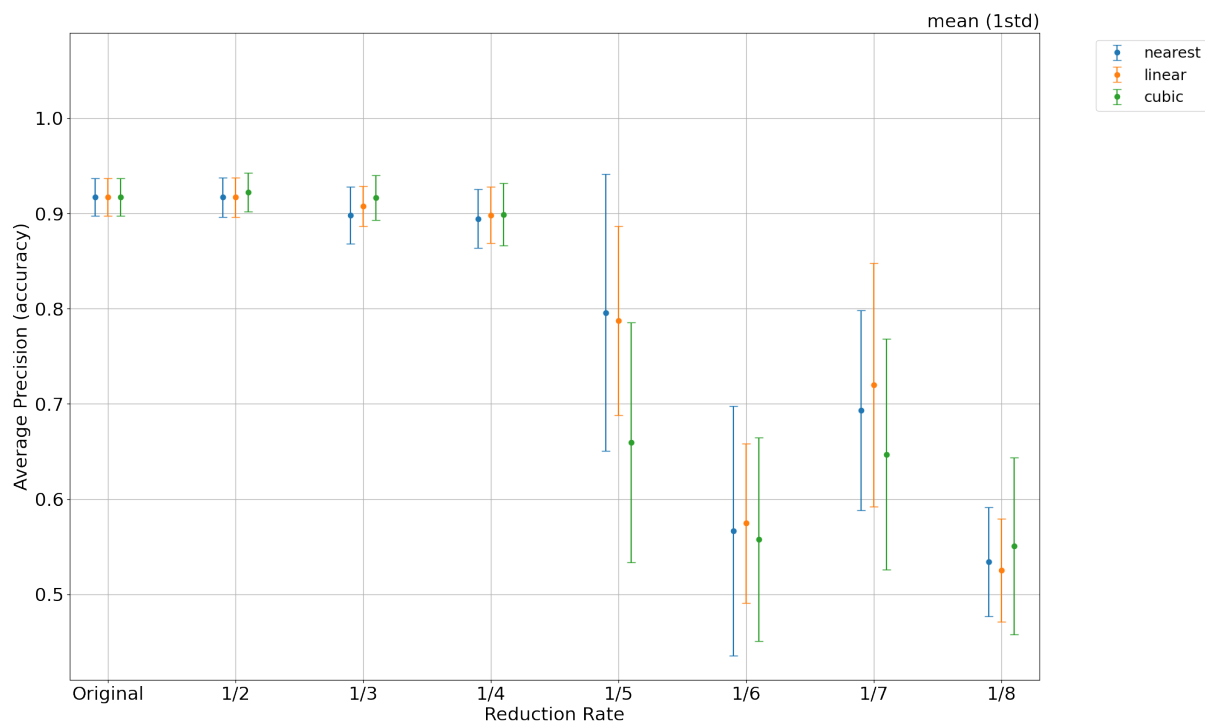


図 5.8: 内挿方法間での比較 (1std)

結果

縮小データの縮小倍率における結果 図 5.5, 図 5.6, 図 5.7 より, Original の結果 (学習データとテストデータの解像度が揃っている状況での予測結果) と同程度かつ accuracy が 0.9 付近を記録しているテストデータの縮小倍率が存在することが分かる.

図 5.5 より, 最近傍補間によって拡大したテストデータに対する予測は, 縮小倍率が $1/2$ から $1/4$ 倍の縮小データに対しては使用可能であることが分かる. なお, 縮小倍率が $1/5$ 倍の縮小データに対する予測においては, accuracy が 0.9 を記録するモデルは 1 標準偏差区間において上位約 17% 程しかなく, モデルの学習にかなりの上振れが起こらないと達成できないスコアであることから, 使用可能ではないとの判断を行った.

同様に, 図 5.6 よりバイリニア補間によって拡大したテストデータに対する予測は, 縮小倍率が $1/2$ から $1/4$ 倍の縮小データに対しては使用可能であり, また図 5.7 よりバイキュービック補間によって拡大したテストデータに対する予測は, 縮小倍率が $1/2$ から $1/4$ 倍の縮小データに対しては使用可能であることが分かる.

このことから, 内挿方法によらず, 学習データとテストデータの解像度差が $1/4$ 倍までであったら, テストデータとして使用可能であることが分かった.

内挿方法間での比較 内挿方法の間で, テストデータに対する accuracy の細かい挙動を比較する. 図 5.8 より, $1/2$ から $1/4$ 倍, および $1/6$ から $1/8$ 倍の縮小データにおいては, 1 標準偏差の誤差区間において内挿方法による有意差は見受けられない. また, $1/5$ 倍の縮小倍率においてはバイキュービック補間によって内挿したテストデータに対する予測の accuracy が他 2 手法より劣って見えるが, 1 標準偏差の誤差区間においては他 2 手法との有意差は見られない. このことから, テストデータへの予測の accuracy において, テストデータの内挿方法における有意差はないことが分かった.

標準誤差 (1std) における予測結果 図 5.5 から図 5.7 では, 標準偏差 (1std) を示した. ここで, 標準誤差 (1ste) の結果を図 5.9 から図 5.11 に示す.

1 標準偏差 (図 5.5 から図 5.7) と比べると, 1 標準誤差の誤差区間では各縮小倍率間で有意差が生じる. 例として, 図 5.9 を見ると, 1 標準偏差の誤差区間では有意差がなかった $1/2$ 倍と $1/3$ 倍の縮小データに対する accuracy に, 有意差が生じていることが読み取れる. このことから, 深層学習において 1 つ 1 つのモデル学習にかなりのブレがあることがわかる.

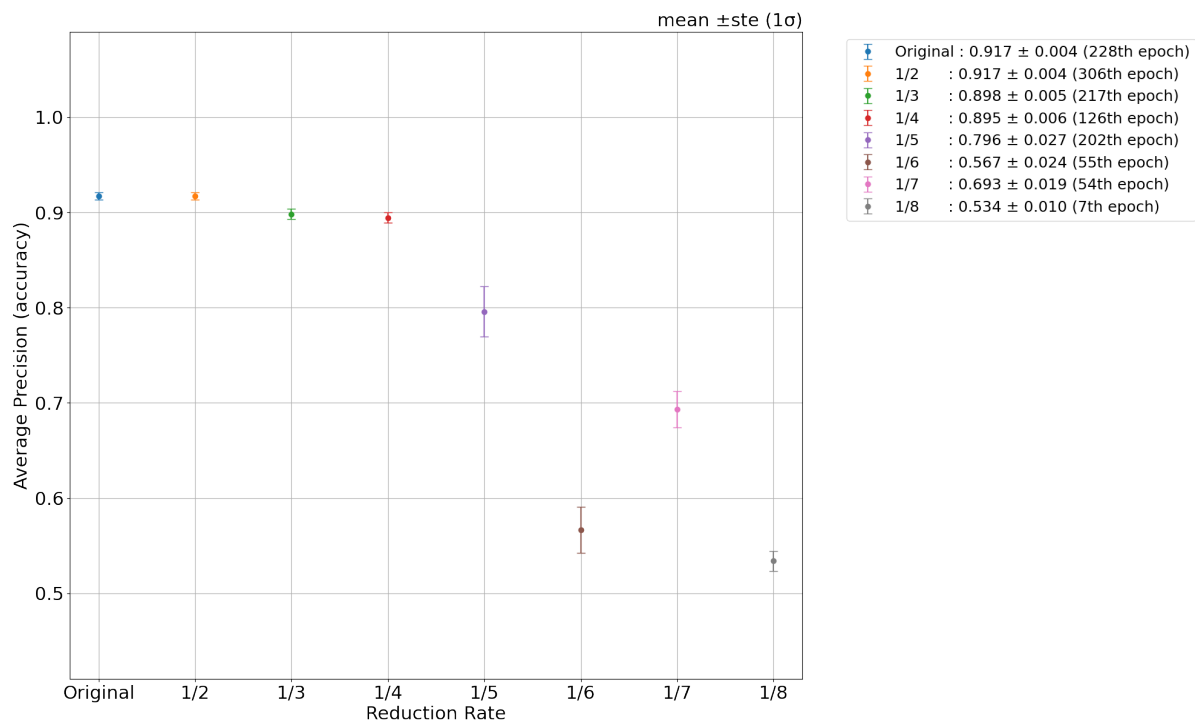


図 5.9: 最近傍補間によって内挿された, 縮小データに対するモデルの予測結果 (1ste)

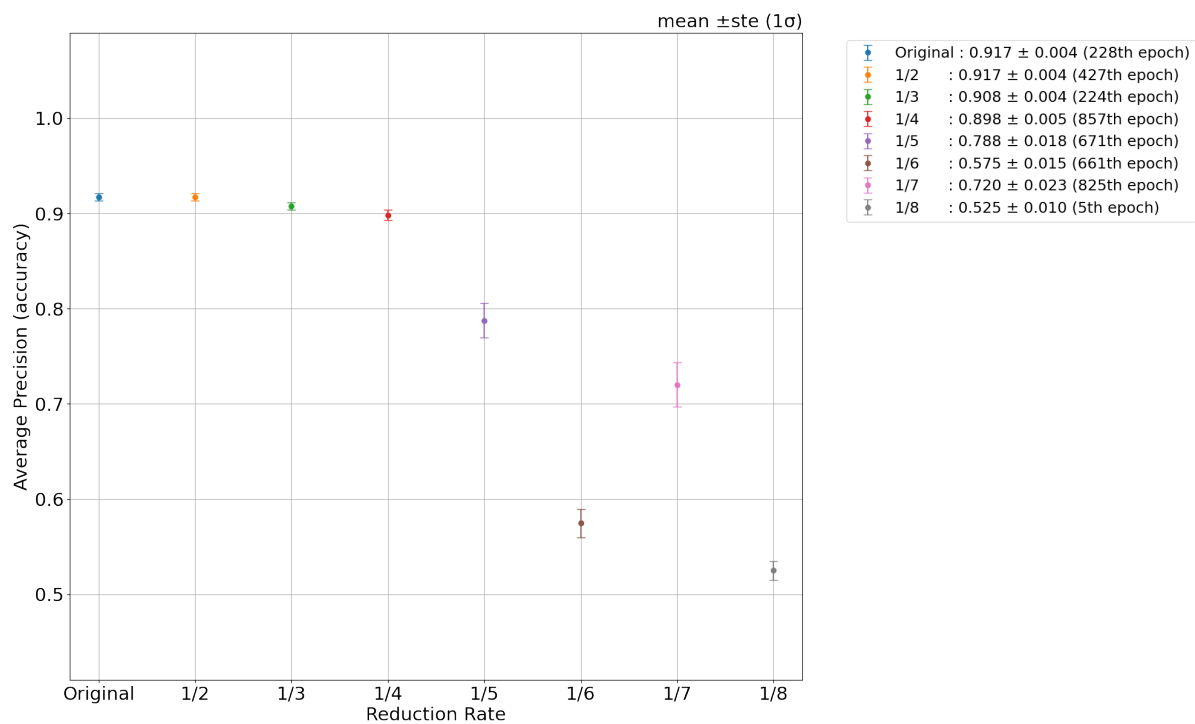


図 5.10: バイリニア補間によって内挿された, 縮小データに対するモデルの予測結果 (1ste)

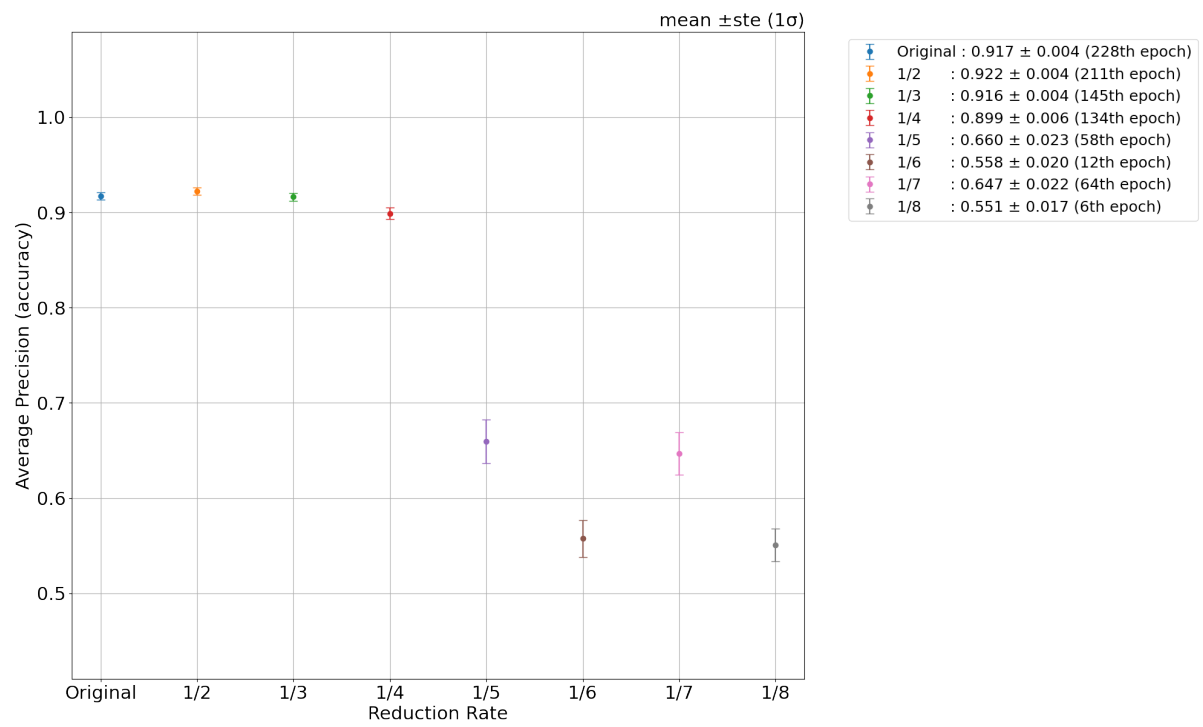


図 5.11: バイキュービック補間によって内挿された, 縮小データに対するモデルの予測結果 (1ste)

5.3 分類精度の使用画像枚数依存性

当論文で掲げている将来展望として、予測精度を高く保ったまま、学習データの枚数をなるべく少なくする方が好ましい。理由としては、天体画像データセットおよびそれをもとにした形態分類カタログの完成には膨大な時間がかかるからである。

今度登場するであろう高解像度データセットだが、完成には時間がかかる。そのため、完成を待たずある程度の高解像度天体データが撮影された時点で、それらを用い高精度の分類モデルを学習し、既存の低解像度データセットに対し形態分類の予測を適用することが将来展望の構想である。そのため、形態分類の精度を高く保ったまま、どのくらいの使用天体数の少なさなら使えるかを検証することは、将来展望を議論するのに有用である。

これを検証するため、分類精度の使用画像枚数依存性について調べる。具体的には、銀河種ごとの使用天体数をそれぞれ 500, 800, 1000 天体とし、モデルの学習・テストを行う。5.2 節で行っていた 300 天体の結果も合わせて、使用天体数 300, 500, 800, 1000 天体で行った実験におけるモデルの予測結果を、テストデータに対する予測の accuracy をもって比較し、以下を検証する。

- テストデータに対する予測の accuracy の、平均値・誤差 (1 標準偏差) の増減はあるのか、増減幅はどのくらいか
- "Original" の結果 (学習データとテストデータの解像度が揃っている状況での予測結果) と同程度の高精度分類を達成するテストデータの縮小倍率に、変化があるのか

5.3.1 実験条件

モデルの学習・テストには、GZ による形態分類が行われた 15,000 天体 (図 4.1 参照) を母集団とし、銀河種類ごとにそれぞれ 500, 800 および 1000 天体を使用した。銀河種類ごとのデータインバランスが起これないようにするため、渦巻銀河と楕円銀河の比率は等しくなるように選定をした。選定を行う際、赤方偏移 z について、 $0 < z < 0.2$ という条件を設定し、あてはまらない天体は除外を行う。これは、地球との距離が遠すぎる天体については上手く特徴抽出が行えず学習の妨げになる可能性があるからである。

使用天体を選定した後、学習データとテストデータの比率が 7:3 となるように天体群を分けた。このうち、学習データには SDSS の切り出し画像、テストデータには SDSS の切り出し画像を縮小した画像をそれぞれ用いた。また、学習データとテストデータ内において、銀河種のデータインバランスが起これないようにするため、渦巻銀河と楕円銀河の比率が等しくなるようにした。

モデルの評価を行う際、モデルの学習およびテストを 30 回行い、accuracy の平均値、標準偏差および標準誤差を導出した。なお、30 回の学習およびテストの際、学習実行毎に使用される天体は毎回シャッフルされる。

モデルの学習期間は 900epoch とした。これは、すべての縮小データの縮小倍率にて、モデルが学習しきる期間、すなわちテストデータに対する accuracy が上がりきる epoch 数が 900epoch であったからである。

5.3.2 実験結果

縮小データに対する予測結果

銀河切り出し画像を $1/2$ 倍から $1/8$ 倍に縮小した縮小データに対する、形態分類モデルの予測結果の図を図 5.12 から図 5.14 に示す。

- 横軸が、テストデータである縮小データの縮小倍率であり、縦軸がテストデータに対する予測の accuracy となっている。
 - － テストデータに対する予測の accuracy について、30 回実行における平均値と標準偏差 (2std) をエラーバー形式で示している。
 - － なお、一番左に位置している Original となっている箇所は、テストデータに縮小データでなく、学習データと同じ大きさの銀河切り出し画像を使用してテストを行った結果である。すなわち、Original が学習データとテストデータの解像度が揃っている状況での予測結果に対し、 $1/2$ から $1/8$ が学習データとテストデータとの間に解像度差がある状況での予測結果を示している。
- accuracy を採る epoch 数は、モデルが十分に学習しきる epoch 数とした。すなわち、テストデータに対する accuracy が最大値となる epoch 数とした。

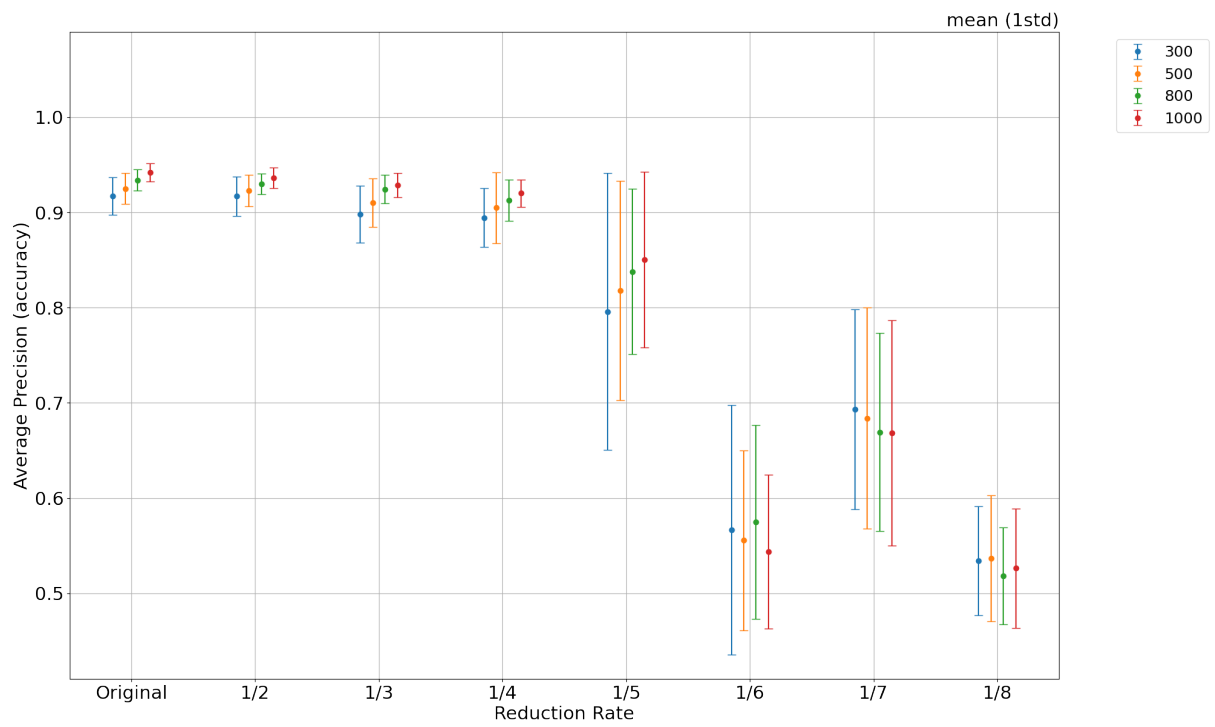


図 5.12: 最近傍補間によって内挿された、縮小データに対するモデルの予測結果 (1std)

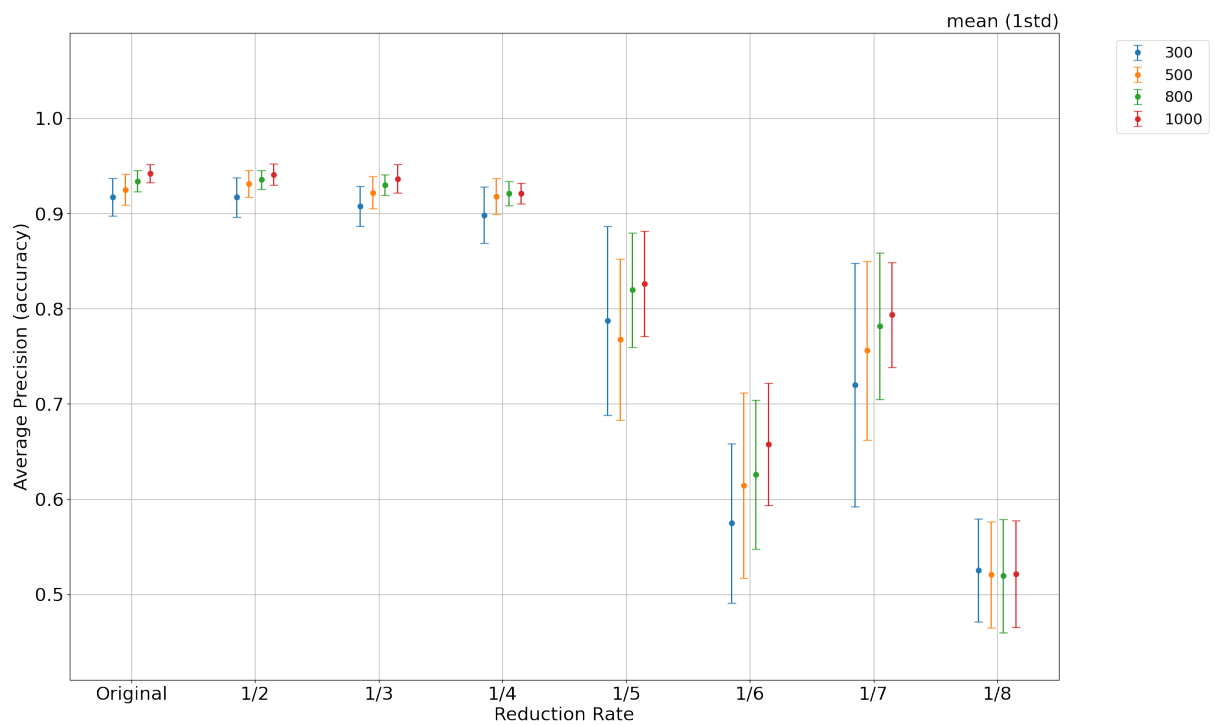


図 5.13: バイリニア補間によって内挿された、縮小データに対するモデルの予測結果 (1std)

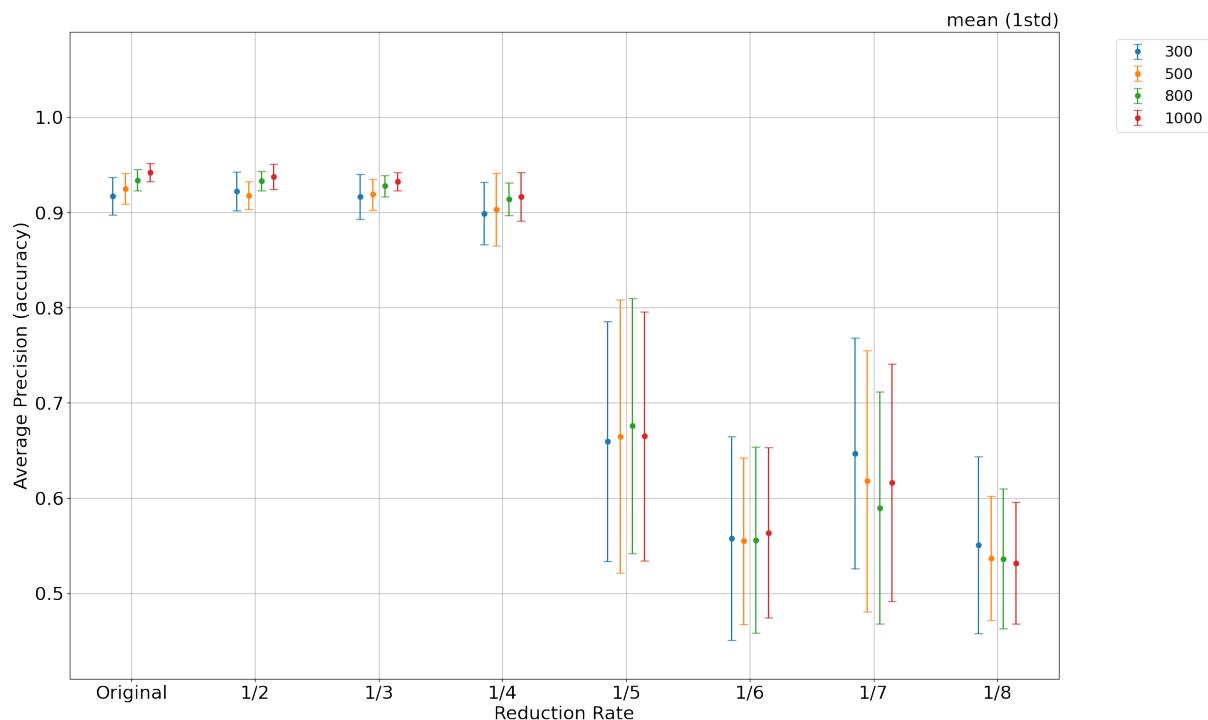


図 5.14: バイキュービック補間によって内挿された、縮小データに対するモデルの予測結果 (1std)

結果

使用天体数増加による accuracy の変化 使用天体数を増やすことで補間方法によらず、縮小倍率が $1/2$ 倍から $1/4$ 倍の縮小画像データに対する予測の accuracy の平均値が上昇し、また標準偏差の値も小さくなった。特に使用枚数が 300 枚と 1000 枚の結果を比較すると、標準偏差の値がおおよそ半減していることが図 5.12 から図 5.14 より読み取れる。

また縮小倍率が $1/5$ 倍から $1/8$ 倍に対する予測の accuracy については、内挿方法により accuracy の挙動に差が出た。これについては**内挿方法間での比較**にて詳述する。

内挿方法間での比較 内挿方法の間で、銀河種ごとの天体使用数を増加させた際に見受けられる、テストデータに対する accuracy の細かい挙動を比較する。なお、縮小倍率 $1/2$ 倍から $1/4$ 倍のデータに対する予測結果は前段落である**使用天体数増加による accuracy の変化**で述べたので、ここからは縮小倍率が $1/5$ 倍以上のデータに対する結果を比較する。

最近傍補間によってテストデータの内挿を行った場合、銀河種ごとの天体使用数を増加させることで、縮小倍率 $1/5$ 倍のデータに対する予測の accuracy の平均値上昇、および標準偏差の減少が見受けられる。しかしながら、標準偏差の減少量は $1/2$ 倍から $1/4$ 倍の縮小倍率データに対する予測と比べるとそれほど大きくないことが分かる。

バイリニア補間によってテストデータの内挿を行った場合、銀河種ごとの天体使用数を 1000 天体まで増やしても、1/5 倍から 1/8 倍の縮小倍率データに対する予測の accuracy が 0.9 を超えず、高精度分類を行えなかった。しかしながら、銀河種ごとの天体使用数を増加させることで、縮小倍率が 1/5 倍から 1/7 倍のテストデータに対する予測の accuracy の精度改善が見込めた。

なお 1/8 倍の縮小倍率に対する形態分類は、銀河種ごとの天体使用数および内挿方法を問わず accuracy 0.5 付近を記録した。このため、学習データとテストデータとの間の解像度差が 1/8 倍である場合、深層学習モデルによる形態予測が行えないのではないと思われる。

バイキュービック補間によってテストデータの内挿を行った場合、銀河種ごとの天体使用数を 1000 天体まで増やしても、1/5 倍から 1/8 倍の縮小倍率データに対する予測の accuracy が 0.9 を超えず、また accuracy の平均値および標準偏差に改善が見られなかった。

使用可能なデータセット間の解像度差について 図 5.12 から図 5.14 より、内挿方法を問わず、銀河種ごとの天体使用数が何天体であっても、縮小倍率が 1/4 倍までの縮小データに対する予測の accuracy が、Original の結果（学習データとテストデータの解像度が揃っている状況での予測結果）と同程度かつ accuracy が 0.9 付近を記録することが読み取れる。このことから、内挿方法・銀河種ごとの使用天体数によらず、学習データとテストデータの解像度差が 1/4 までならば使用可能であることが分かった。

また**内挿方法間での比較**にて、縮小倍率が 1/4 倍までの縮小データに対する予測では、銀河種ごとの天体使用数を増加させることで内挿方法によらず予測精度が向上することが分かった。このことから、学習データとテストデータの解像度差が 1/2 倍から 1/4 倍ほどである 2 値銀河形態分類は、モデルの学習の上振れを期待せずとも、高精度の形態予測を期待できる分類問題であると思われる。

次に、学習データとテストデータの解像度差が 1/5 倍より大きい場合について述べる。最近傍補間によってテストデータを内挿する場合、解像度差が 1/5 倍のとき、銀河種ごとの天体使用数を増やすことで誤差区間が減少する。しかし、最も多い天体数である使用天体数 1000 天体においても、1 標準偏差区間の上位 25%しか accuracy が 0.9 を超えず、このことから、モデルの学習が上振れることを期待する必要があるため使用不可な解像度差であると判断した。

また、図 5.13、図 5.14 より、バイリニア補間、およびバイキュービック補間においては、Original の結果と同程度かつ accuracy が 0.9 付近を記録しないことが読み取れる。

したがって、学習データとの解像度差が 1/5 倍より大きいテストデータは使用不可であることがわかった。

5.4 5章全体の議論・結論

どの補間フィルタを用いるべきか

第5章では学習データに高解像度、テストデータに低解像度の画像データを使用していた。深層学習モデルに入力する画像データは全て同じサイズである必要があるため、今回の実験ではテストデータである低解像度画像を拡大を行った。

ここで、今回の画像拡大に用いた3手法(最近傍補間・バイリニア補間・バイキュービック補間)のうち、どれを使うべきかを考察する。

まず、第5章にて”使用可能”と定義した、学習データとテストデータとの間の解像度差が1/2倍から1/4倍の状況でモデルの学習・それによる予測を行う場合について考察する。図5.12から図5.14より、どの天体使用数においても、3つの拡大手法での accuracy は同程度の挙動、なおかつ0.9付近のスコアを記録している。このことから、学習データとテストデータとの間の解像度差が1/2倍から1/4倍の状況でモデルの学習・それによる予測を行う場合、これら3手法のどれを使用しても高精度の予測が行えと考えられる。

ただし図5.14より、バイキュービック補間は使用天体数を増やした際の accuracy 平均値向上・誤差低下の恩恵を受けづらいことが読み取れる。

次に、第5章にて”使用不可”と定義した、学習データとテストデータとの間の解像度差が1/5倍から1/8倍の状況でモデルの学習・それによる予測を行う場合について考察する。この場合、5.3.2の結果における**内挿方法間での比較**より、使用天体数を増加させた際の accuracy 平均値向上・誤差低下が最も見込めるバイリニア補間が適していると思われる。

最後に、3つの拡大手法の処理にかかる時間を考える。時間計測を行うため、SDSSから取得した銀河切り出し画像を1/8倍に縮小した縮小画像を、GZの分類ラベルにてラベル付けされた渦巻銀河50天体・楕円銀河50天体について拡大処理を実行した。拡大処理は各手法にてそれぞれ10000回ずつを7周実行し、拡大にかかった時間の平均値および1標準偏差を導出した。

表5.1に3つの拡大手法の処理にかかる時間を示す。表より、最近傍補間が最も処理が軽く、バイリニア補間が2番目、バイキュービック補間が3番目に処理に時間がかかることが読み取れる。

上で述べた3つを総合的に考慮すると、モデルの学習・テストにおける処理を最も軽くしたい場合は最近傍補間が望ましいと考えられる。

また処理が少し重くてもよく、使用天体数を多くした際の accuracy に対する恩恵を受けたい場合はバイリニア補間による内挿がよいと思われる。

バイキュービック補間は、使用天体数を増加させた際の accuracy に対する恩恵を最も受けづらく、また拡大処理における処理の重さを考えると、3手法の中では使用する意味が最も薄いと考えら

れる。

表 5.1: 各補間フィルタによる拡大の計算コスト (平均値 (1 標準偏差))

最近傍補間	$247\mu s \pm 2.25\mu s$
バイリニア補間	$378\mu s \pm 5.37\mu s$
バイキュービック補間	$1.48ms \pm 7.74\mu s$

学習データとテストデータとの間の解像度差が 1/6 および 1/7 の際の、テストデータに対する予測 accuracy の挙動

第 5 章にて使用天体数・内挿方法を問わず共通の傾向であった、「縮小倍率が 1/7 倍の縮小データに対する予測精度が、縮小倍率 1/6 倍の予測精度より高い」事に対し、考察を行う。

一般的に、テストデータの解像度が低ければ低いほど、分類モデルによる予測精度は低くなると考えられる。図 5.12 から図 5.14 においても、縮小倍率が 1/2 倍から 1/6 倍においては上記の傾向が見受けられる。しかしながら、1/7 倍の縮小倍率に対する予測精度は、1/2 倍から 1/6 倍にて見受けられた傾向から逸脱している。この傾向に対し、考察を行う。

1/6 倍、1/7 倍の縮小画像に対する画像拡大例を図 5.15 に示す。図 5.15 より、拡大手法によって生成される画像が、見かけではかなり異なることが分かる。しかし、図 5.12, 図 5.13 を比較すると、拡大手法によって縮小倍率 1/6 倍と縮小倍率 1/7 倍の accuracy の挙動に大きな差はないことが分かる。ここで、1/6 倍、1/7 倍の縮小画像 (図 5.15 の左下) を見ると、画像内の最も明るいピクセルが存在する場所に差があることが分かる。1/6 倍の縮小画像では中央より右下、1/7 倍の縮小画像では中央が一番明るいピクセルであることが読み取れる。このことから、平均画素法による縮小の際、何らかの形で銀河の形態的特徴が損失している可能性がある。そしてこの形態的特徴損失により、学習データとテストデータとの間の解像度差が 1/6 および 1/7 の際の、テストデータに対する予測 accuracy の挙動に影響を及ぼしていると思われる。

使用枚数が何枚までならば高精度を保てるか

最後に、モデルの学習・テストに用いる銀河種ごとの天体使用数について、形態分類の精度を高く保ったまま、どのくらいの使用天体数の少なさなら”使用可能” かについて議論を行う。なお、第 5 章全体を通して”使用可能” と定義している学習データとテストデータとの間の解像度差が 1/2 倍から 1/4 倍の状況に限定して、使用天体数の議論を行う。

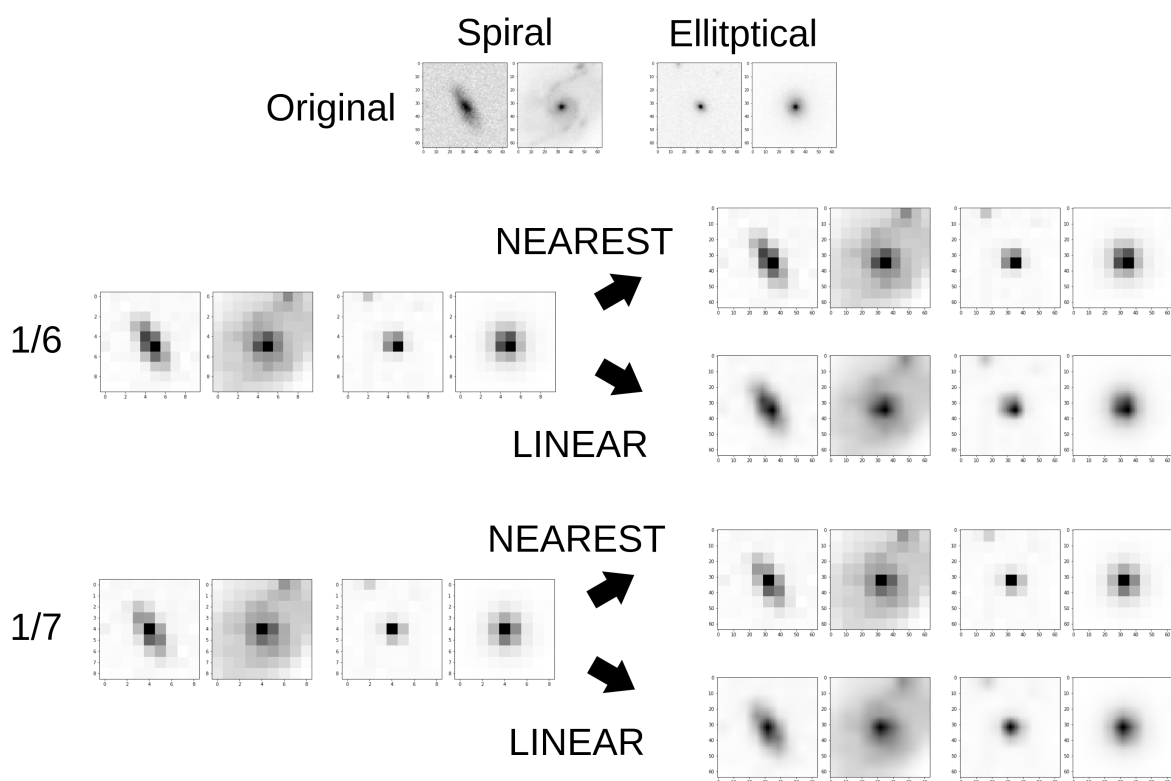


図 5.15: 1/6 倍, 1/7 倍の縮小画像に対する画像拡大例 (最近傍補間・バイリニア補間)

図 5.12 から図 5.12 より, 使用する天体数は何天体でもよいと思われる. これは, 内挿方法を問わず, "Original" と同程度の予測精度, すなわち accuracy が 0.9 程度をスコアするからである. ここでテストデータに対する予測の accuracy を 0.9 以上にしたい場合だが, 図 5.12 から図 5.12 より, これは銀河種ごとの天体使用数を 800 天体以上にすることが望ましいと考えられる.

この章では, 「学習データとテストデータとの間の解像度差がある状況において, 分類モデルの学習が行えるか」の検証を行うため, 2 つの実験を行った.

学習データとテストデータとの間の解像度差が 1/4 までならば, 内挿方法によらず使用可能, すなわち学習データとテストデータの解像度が揃っている状況での予測結果である "Original" と同程度の高精度な分類, すなわち accuracy が 0.9 付近を記録する結果となった.

銀河種ごとの使用天体数を増加させた際, テストデータに対する予測の accuracy の挙動は内挿方法によって差が生じた. 最近傍補間による内挿では, 1/2 倍から 1/5 倍までの縮小倍率画像に対し精度改善が見込めた. バイリニア補間に関しては, 1/2 倍から 1/7 倍までの縮小倍率画像に対し精度改善が見込めた. バイキュービック補間は 1/2 倍から 1/4 倍までの縮小倍率ならば精度改善が見込めたが, 他 2 手法よりもその恩恵は小さく, 更に 1/5 倍より大きい縮小倍率画像に対する予測精度は改善が見込めなかった.

内挿方法によってテストデータに対する予測精度が異なる挙動を見せることから、モデルへの内挿方法を模索することで、更なる精度改善、および当論文では使用不可と判断した学習データとテストデータとの間の解像度差においても、高精度分類が行える可能性がある。

第6章 議論

6.1 今回の実験から得た結論

当論文の将来展望は、高空間分解能観測装置データを用いてモデル学習を行うことで、既存の低空間分解能データセットに対し更なる高精度形態分類を提供するというものである。この将来展望が実現できる可能性を検証するため、当論文では第4章、第5章にて将来展望の前段階となる実験群を行った。その結果、深層学習による銀河形態分類の主に渦巻銀河・楕円銀河に分類する2値分類問題において、学習データとテストデータとの間に1/4倍の画像解像度差が生じていても、テストデータに対し accuracy が0.9付近となる高精度分類が提供できることが分かった。このことから、将来展望である「異なる観測装置データの組み合わせを行い、深層学習による銀河形態分類」を実現できる可能性が示された。

一方で、SDSS DR7とGZ1を組み合わせた深層学習による形態分類の更なる精度向上など、多くの将来課題が残されている。6.2節ではそれらについて言及を行う。

6.2 将来課題

学習した分類モデルの精度のよさについて

モデルの分類精度を向上するべきである。

将来展望（ポンチ絵）は、「高→低」が「低→低」を上回る。

（その際、「高→低」は「高→高」レベルの分類精度であることが望ましい。）

それはモデルを学習させるデータがより細かい特徴を持った高解像度画像だからって理由。

しかし、現状では「低→低」の組み合わせでも割と分類でき、「高→低」に匹敵する分類精度を持っている。

「低→低」で割と行けるなら、「高→低」や「高→高」でももっと行けるはず。

精度向上案

GZ の論文（どれか忘れた）や cheng+(2019) より,「いろんな色を組み合わせると, 形態的特徴が増える」

→ 用いる band 数増やしてみる

モデルのハイパパラメータチューニングを行う.

今回のモデルは SDSS の解像度など, 今回の実験条件に最適化されておらず, モデルのチューニングを行うことで精度向上が見込める可能性がある.

実際に未知天体データセット (ラベル付けが為されていないデータセット) に用いる時の方法

どうやって最高のモデルを残すの?

→ 今回みたいに〇〇回実行したあと, その地点までのモデルを k 分割交差検証法?

何個もモデル作って未知天体に適用. 一番多いラベルを予測結果とする という使用方法

GZ1 の不確かさを, 深層学習でどう扱うか

・今回は GZ1 の不確かさについて注目しているが, 他にも人間の目による分類がまだ提供されていないデータセット (ex. HSC の銀河データセット) に対し先んじて提供, 等の使い道があるかもしれない

教師なし学習利用

GZ1 の分類 (人間の目による分類) と比較して, 教師なし学習による分類がどれくらい信頼できるかを調べれば, GZ1 に不確かさと分類された天体に渦巻銀河・楕円銀河の 2 値分類を提供できるかもしれない

6.3 将来展望

当論文にて異なる観測装置データの組み合わせによる深層学習形態分類の可能性が示された. そこで, ここからは実際に異なる観測装置データの組み合わせを行った結果を示す.

6.3.1 実験概要

学習データである高解像度データセットとして, HSC から取得した銀河切り出し画像を用いた. ここで, HSC とは... (HSC の説明)

学習データの正解ラベルだが、これは HSC の銀河画像を人間の目による分類を行った分類ラベルが存在しなかったため、GZ1 による分類ラベルを流用した。

テストデータである低解像度データセットとして、第 4,5 章にて用いてきた SDSS から取得した銀河切り出し画像を使用した。ここで、HSC の空間分解能が $0.168''/\text{pix}$ 、SDSS の空間分解能は $0.369''/\text{pix}$ であり、これらの分解能の差はおよそ 2.3 倍である。これは第 5 章にて結論付けた”使用可能”である解像度差の範囲内である。

モデルの学習・テストに使用する天体数として、銀河種ごとに 1000 天体ずつを使用した。選定された天体は、学習データとテストデータの比率が 7:3 となるように振り分けられた。なお、学習データおよびテストデータ内における渦巻銀河と楕円銀河の比率は等しくなるようにした。取得を行う際、赤方偏移 z について、 $0 < z < 0.2$ という条件を設定し、あてはまらない天体は除外を行う。これは、地球との距離が遠すぎる天体については上手く特徴抽出が行えず学習の妨げになる可能性があるからである。

今回用いた深層学習モデルは、cheng et al.(2019)??にて用いられていた銀河形態分類モデルを参考にした。今回用いたモデルの構造を図 6.1 に示す。このモデルは畳み込み層を合計 3 つ有しており、それぞれのカーネルサイズは 3×3 , 3×3 , 2×2 である。それぞれの畳み込み層の後には、 2×2 の max-pooling 層が存在する。全畳み込み層の後に全結合層が 2 層配置されており、それぞれ 1024 個のノードを有している。

モデルの評価を行う際、モデルの学習およびテストを 30 回行い、accuracy の平均値、標準偏差および標準誤差を導出した。なお、30 回の学習およびテストの際、学習実行毎に取得される天体は毎回シャッフルされる。

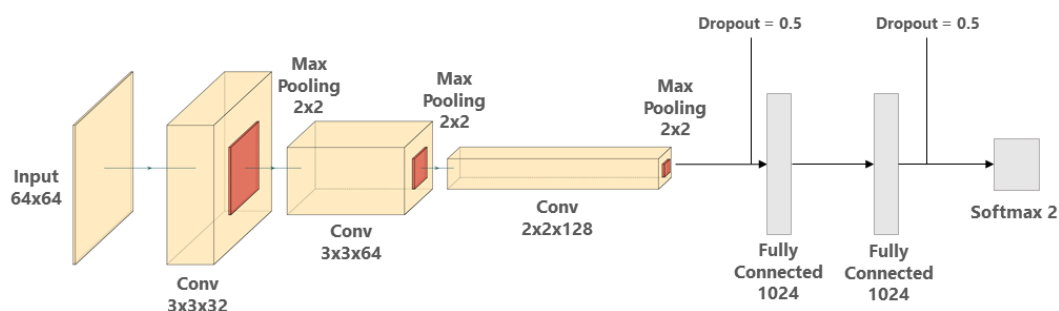


図 6.1: 用いた分類モデルの構造図

6.3.2 実験結果

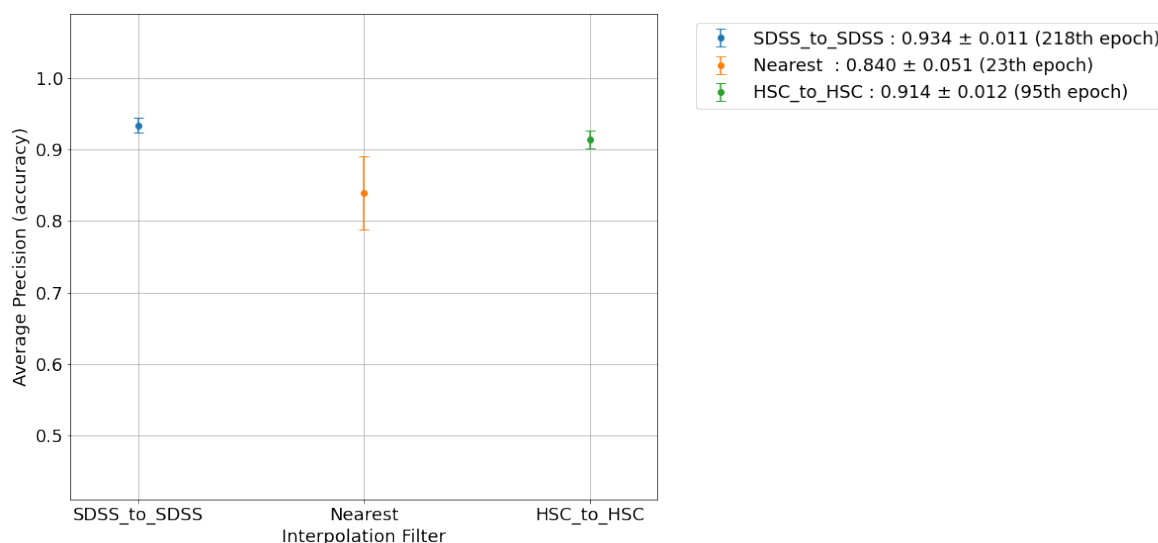


図 6.2: HSC で学習, SDSS に対しテストを行った結果 (accuracy の平均値 (1 標準偏差))

6.3.3 結果

hsc→hsc の予測精度より, sdss→sdss の予測精度の方が高い. すなわち, 高解像度データによって学習したモデルの予測より, 低解像度データで学習したモデルの予測の方が結果が良くなっている.

この原因として, 2つの原因が考えられる.

- ・GZ1 の形態ラベル付けが間違っている. 例として, 渦巻銀河に対し楕円銀河というラベルが GZ1 でつけられたとする. SDSS の解像度では楕円銀河に見えるが, 高解像度である HSC では渦巻銀河にしか見えないとき, モデルの学習に支障が生じる.
- ・テストデータへの汎化性能が高く, 高解像度画像データから複雑な特徴を学習できるモデルを用いらないと, 精度が低くなる可能性.

第7章 おわりに

謝辞

本研究を進めるにあたり、ご指導を頂いた飯田佑輔准教授および東京理科大学の大井渚様、そしてデータセット作成や宇宙関連知識取得に際し甚大な助力をいただいた同研究室の津田様に、厚く感謝申し上げます。

また、日常の議論を通じて多くの知識や示唆を頂いた飯田佑輔研究室の皆様に感謝いたします。

SDSS および SDSS-II の資金はアルフレッド・P・スローン財団から提供され、また参加機関は米国科学財団、米国エネルギー省、米国航空宇宙局、日本の文部科学省、マックスプランク協会、英国高等教育基金協会です。SDSS の Web サイトは、<http://www.sdss.org/> です。

SDSS は参加機関のための天体物理学研究コンソーシアムによって運営されています。参加機関は、アメリカ自然史博物館、ポツダム天体物理学研究所、バーゼル大学、ケンブリッジ大学、ケース・ウェスタン・リザーブ大学、シカゴ大学、ドレクセル大学、フェルミラボ社、高等研究所、日本参加グループ、ジョンズ・ホプキンス大学、原子核宇宙物理学合同研究所、カブリ粒子宇宙物理学研究所、韓国科学者グループ、中国科学者グループ、中国科学者グループ、韓国科学者グループ、韓国科学者グループ、中国科学院 (LAMOST)、ロスアラモス国立研究所、マックスプランク天文学研究所、マックスプランク天体物理学研究所、ニューメキシコ州立大学、オハイオ州立大学、ピッツバーグ大学、ポーツマス大学、プリンストン大学、米国海軍天文台、ワシントン大学です。