修士論文

敵対的生成ネットワークによる講義アーカイブシステムの板書鮮明化

1810074　CUI BOWEN

主指導教員　長谷川　忍

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和３年３月

Abstract

In recent years, with the process of video recording, speeding up the Internet, and the capacity of data storage devices increasing, we have entered an era in which the captured video can be stored in the cloud and rebroadcast as well as live on the Internet. In 2017, the domestic Internet usage rate reached the 80% level at 80.9%, and by terminal, the usage rate on smartphones was 59.7%, which exceeded personal computers (52.5%). In addition, a network that can distribute and receive high-quality video content has been established, and an infrastructure that allows users to enjoy various video distribution services has been established. Furthermore, "hobbies / practical use / education" (30%) was ranked after "music" (45%) as a genre to be viewed on the video distribution service. From these facts, it can be said that video distribution using networks has become familiar to us.

Since the 1990s, e-learning, which means computer-based learning and education, has attracted attention. With the spread of the Internet, e-learning has become easier to realize. Also its educational value is drawing attention. Such e-learning system teaching materials are in a multimedia form that combines video, audio, board writing, and power-point. The users are supposed to be "learners" and "teachers". However, the traditional relationship between learners and teachers has changed. Learners can study in their free time and at any place. However, it is difficult to communicate in Two-way, such as ask questions. Teachers can work efficiently, but it is difficult to grasp the situation of learners. The lecture archive system is one of the e-learning systems. It is also an approach that records the video and audio of lectures held in the lecture room as digital data and used as the content of the asynchronous remote learning system.

In this research, we focused on the problem that the whiteboard board is difficult to read in the lecture archive system introduced in higher education institutions. The hardware solution proposed in the previous research takes time and cost to introduce new equipment to all rooms. It is not realistic because when delivering high-definition video such as 4K, the network bandwidth becomes very large. Also, the existed lecture video in the system cannot be clarified. JAIST has introduced a lecture archive system since 2006, and there are already many recorded videos, the sharpening method on software is more meaningful.

In addition, there are several super-resolution processing software methods that are applicable to lecture archive systems. In recent years, deep learning techniques such as SRCNN and SRGAN have become popular and have shown sufficient effects as a general image super-resolution processing method. However, the same effect cannot always be obtained, even in writing on a whiteboard. Furthermore, the characteristic of the lecture is that the changes in the writing on the board are not drastic. Compared to movies and animations, the written characters remain for a long time. Therefore, as the target of super-resolution processing, we targeted screenshots of the board writing area extracted from the videos. In this research, we expect when the SRGAN processing is applied with the expectation, a sharper image could be generated, and a high sharpening effect would be obtained.

In this research, when applying SRGAN to the lecture archive system, some improvement measures were applied. As a model, the Subpixel layer, Convolution layer, ReLU layer were deleted for 2x upscaling factors. Sigmoid layer was deleted as the cause of the disappearance of the learning gradient to perform super-resolution processing. Furthermore, to improve the accuracy, a Skip Connection was added to Discriminator. In

performing general preprocessing, we also devised the method improvement which suitable for the lecture archive system. We realized in previous studies, the characters on the board were often binarized. By performing binarization processing, it is possible to increase the characteristics of the surplus on the board easily. However, in the lecture archive system, the character color of the board may be important. For example, by changing the character color, the part that the teacher wants to pay attention to and the part that the teacher does not want to pay attention to is often expressed. To retain such color information, we decided to clarify the color characteristics.

The contrast between the characters and the background is quite low in the hard-to-read board. Therefore, we thought about identifying the characteristics of the board writing and treating the board writing and the margins separately, but it is very difficult to separate them. Furthermore, the actual board writing environment also differs each time due to various factors such as the color of sunlight, the brightness of the classroom, and the color of letters. In other words, such a pretreatment method is by no means highly versatile, and if the shooting environment and the color of the pen used for writing on the board are not restricted. It is necessary to consider other pretreatment methods.

So, we used Contrast Limited Adaptive Histogram Equalization and found effective parameters to writing data. The noise was slightly higher with such adjustments than that of the SRGAN method, but the contrast and color depth could be significantly increased. In the evaluation experiment, the evaluation of readability by the subject exceeded the methods such as SRCNN and BICUBIC, and the best result showed the validity of the proposed improvement measure.

However, GAN-based super-resolution processing may generate new undesired features, although it can produce sharp images. This hinders the reading of small text details. Comparing the proposed method with SRCNN, the CNN-based super-resolution processing method has higher reproducibility, and the detailed features can be almost retained. However, from the detailed objective evaluation of this study, the correct answer rate in all the super-resolution methods is less than 40%. So, it is necessary to consider a better super-resolution processing method for reading characters with a small number of features.

At the end of this research, we have some findings about the future tasks. First, one of them is to improve the recall rate of high-definition board writing with SRGAN. In this proposed method, the SRGAN method process the writing on the board in an easy-to-read manner, but for details, we think that the high reproducibility point of SRCNN can be referred to. In the evaluation experiment, it was confirmed that the CLAHE pretreatment had a good effect. However, it will deteriorate the image quality, generate noise, and was not suitable for subjects who were concerned about noise. Therefore, the second direction is to consider a method that sets noise elimination as a set. Finally, VGG and MSE loss functions are used in SRGAN. The loss generated by comparing the features of the generated image and the original high-resolution image was affected by the margins of the whiteboard area, and it will cause a problem that the learning depth cannot be deepened. It is considered necessary to set an innovation loss function that suitable for the writing on the whiteboard.

Keywords: lecture archive system, whiteboard, super-resolution, GAN, CLAHE.

目次

図目次

表目次

第1章 はじめに

近年ビデオ撮影とネットワークの高速化とデータ保存デバイスの大容量化に伴い，撮影したビデオをクラウドに保存し，ネットで再放送することだけでなく，生放送もできる時代となった．2017 年の国内のインターネット利用率は 80.9%と 8 割台に達し，端末別ではスマートフォンでの利用率が 59.7%とパソコン（52.5%）を上回った[24]．また，高品質な映像コンテンツを配信・受信可能なネットワークが整備され，多様な動画配信サービスを享受できるインフラが整ってきている．さらに，動画配信サービスで視聴するジャンルとして，「音楽」(45%)に続き，「趣味・実用・教育」(30%)がランクインした[24]．これらのことから，ネットワークを活用した動画配信は我々の身近なものとなったといえる．

このような背景のもとで，北陸先端科学技術大学院大学（以下：JAIST）においては，収録された講義が学内ネットワークで配信され，Web ブラウザ上で利用できるサービスとして提供されている．ここでは，講義室で行われる対面講義の映像・音声をデジタルデータとして収録し，体系的に管理・配信するものを講義アーカイブと呼ぶ[5].

吉良(2015)[1]の報告によると，JAIST の講義アーカイブシステムは「講義内容の理解が深まった」の項目の満足度が高く，学生はアーカイブを対面講義の補完的教材として効果的に利用できていると考えられる．また，「これからも講義アーカイブを利用したい」の項目も高い満足度を得ており，講義アーカイブシステムは学生におおむね好評であることがわかった．これらのことから，講義アーカイブは学生の予習や復習に大きな意義のあるシステムであると言える．特に今年度はコロナウイルスの影響で，ほぼ全ての授業をオンラインの形式で行わざるを得ない状況となっている．しかしながら，講義のための主要なリソースとして講義アーカイブシステムを捉えると，様々な課題が挙げられる．図 1.1 に 2017 年に行われた JAIST 講義アーカイブシステムの利用者アンケートの結果[2]を示す.「ホワイトボードの読みやすさ」の平均値は 5 段階評価の 2.6 であり，全体の項目の評価平均値の 3.7 より低かった．また，「ホワイトボードの文字が薄く，読み取れないときが多々あった」,「字がぼやけて読みにくい」という自由記述のコメントも見られた[2]．このように，講義アーカイブにおけるホワイトボードの板書が読みにくいことは講義内容の理解を深めるための支障になっていると言える．

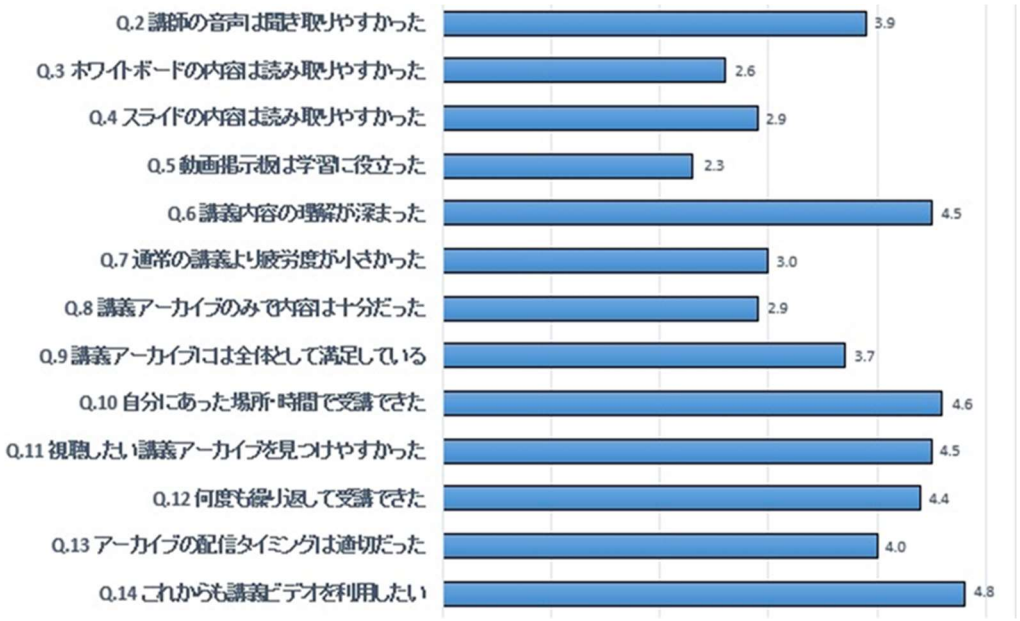


図 1.1：2017 年 JAIST アーカイブシステムの利用者アンケート[2]

本研究の目的は，講義アーカイブにおけるホワイトボード領域の板書を鮮明化することである. 本学の講義アーカイブシステムでは SD 画質のみならず HD 画質も提供されているが，ホワイトボードの細かい文字についてはまだ読み取れない場合が多い．新たな機器を全ての部屋に導入するには時間とコストがかかることや，4K などの高精細な動画を配信するとネットワーク帯域が非常に大きくなるため，本研究では新しいカメラや複数の専用カメラの設置などをせずに，直接収録された動画に超解像処理を行うことを目指す．なお，本研究では教師の位置により隠された板書部分，またホワイトボード領域以外の部分は考慮しないこととする. まとめると，ホワイトボード領域に適用可能な手法により鮮明化を行い，拡大時にシャープな板書を生成することで，アーカイブの読みやすさを追求することが主要な目的である.

本論文は以下のように構成される．2 章では本研究の関連研究として，講義アーカイブで行われた研究や画像処理領域の超解像処理技術などについて述べる. 3 章では講義アーカイブシステムにおける読みにくさの問題を解決するために，本研究の提案手法を詳しく述べる. 4 章では本研究の提案手法の結果とそれに対しての評価実験の結果について示す. 5 章では本論文のまとめと今後の課題を述べる.

第2章 関連研究

2.1 講義アーカイブシステム

1990 年代から，コンピューターを利用した学習や教育，いわゆる e-learning が注目されるようになった．インターネットの普及に伴い，e-learning はより簡単に実現できるようになっており，その教育価値は注目されている．このような e-learning システム教材は動画や音声，板書，ppt などを組み合わせたマルチメディア形態である[25]．利用者としては，「学習者」と「教師」が想定されているが，従来型の学習者と教師の関係とは変化している．学習者は自由な時間と場所で学習できるが，質問などといった双方向でのコミュニケーションをしにくいという問題点がある．教師は効率的に業務できるが，学習者の状況を把握しにくいことが問題であると言える[25]．本研究で対象とする講義アーカイブシステムとは，e-learning システムの一種であり，講義室で行われる対面講義の映像・音声をデジタルデータとして収録し，非同期遠隔学習システムのコンテンツとして活用するアプローチである[3]．

2.1.1 JAIST 講義アーカイブシステム

JAIST では，講義アーカイブシステムが 2006 年度より情報科学研究科の講義において導入されており，2020 年度にはコロナ禍への対応のため，全学の講義で利用された．本研究では，JAIST の講義アーカイブシステムを対象とし，そのシステムで撮影した動画をデータとして研究を行う．本節では，本学の講義アーカイブシステムについて説明し，その特徴と問題点，及び先行の研究方法について議論する．

図 2.1 に本学における講義アーカイブシステムの構成を示す[4]．収録対象の講義室には，アーカイブ収録サブシステムとして，4K 対応ビデオカメラが天井に設置されている．一方，エンコード装置としては，フル HD 品質のエンコードが可能な Photron PowerRec SS が導入されている[5]．全てのエンコード機器はサーバ室に集約されており，各講義室の映像・音声は PC 画面とミキシングを行って講義室とサーバ室の間を接続する光ファイバで転送し，マトリックススイッチャに収容される構成となっている[5]．

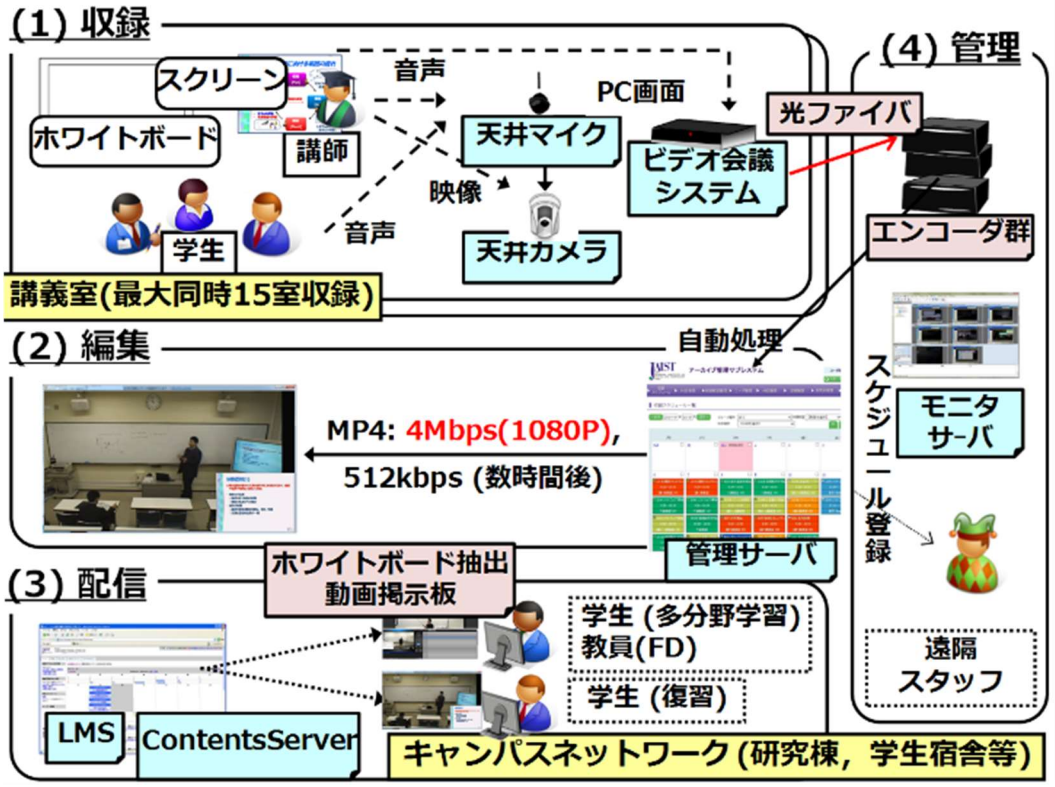


図 2.1：講義アーカイブシステムの概要[4]

2.1.2 講義アーカイブシステムの鮮明化

大西ら（2002）は，カメラワークを自動化することで板書画像の品質を向上させる手法を提案し[6]，市村ら（2006）は，複数台のカメラを用いて板書を保存することで解像度を向上させる手法を提案した[7]．矢田ら（2004）は，2 台のカメラおよびペン位置検出装置を用いて鮮明な板書画像の抽出を実現した[8]．しかしながら，複数台のカメラの設置が必要である[9]．現在様々な大学で導入されている講義アーカイブシステムにこれらの試みを適用するためには，新たな機材の導入が必要となるほか，収録から配信までの自動化処理を実現することが困難である[5]．

Ni（2016）[10]は OpenCV の SuperResolution クラスを使用した超解像手法をホワイトボードへ適用した[10]．スマートフォンで撮影したホワイトボードの板書が不鮮明である問題に対し，複数のフレームを利用した超解像処理を用いていた．その結果，視聴体験がある程度改善され，ホワイトボードへの超解像化適用の有効性が示された[10]．

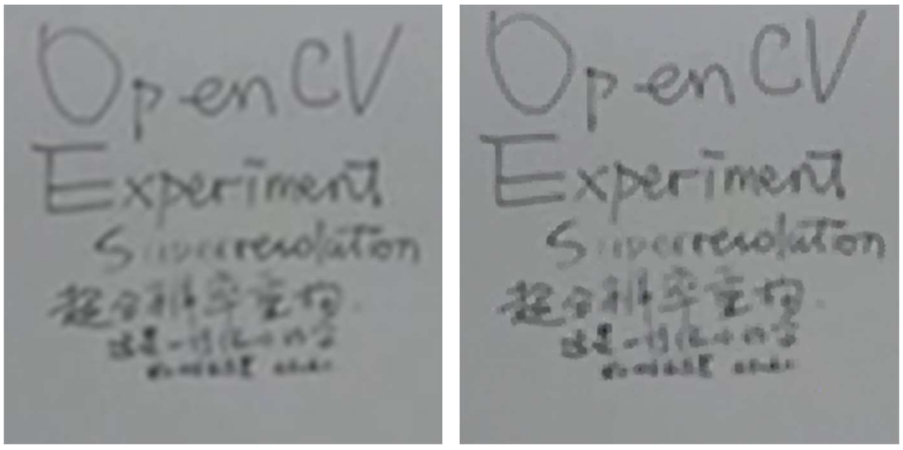


図 2.2：OpenCV の SuperResolution による超解像処理結果[10]
左原本，右超解像処理後

小林（2017）[5]は Ni の研究を踏まえ，講義アーカイブの複数フレームに対して超解像処理を適用可能なシステムを開発した．その中で，二値化によるホワイトボード領域抽出やフレーム間差分，教師領域の検出，教師位置の推定などを実現した[5].

しかし，板書の鮮明化については十分ではなく，ぼやけたり，読みにくい部分多かったりしているのが現状である．また，本研究で対象とするような固定カメラで収録された映像の場合，損失した特徴が少なく復元効果が限定的になるため，複数のフレームから画面の失われた特徴を復元する手法には限界があると考えられる．

そこで，本研究では講義アーカイブの映像そのものではなく，講義映像から切り出した板書の静止画を対象とする．e-learning の中心としての板書に注目し，これまでの超解像化手法より読みやすい超解像画像を生成する手法について検討する．

2.2 超解像手法

現時点でソフトウェアにより画像を拡大し，超解像化する手法としては画素の補間法（Nearest neighbor，Bilinear，Bicubic 法），CNN，GAN などがある．その中で特に講義アーカイブシステムに適用し，効果があると想定される手法について説明する．

2.2.1 画素補間法

画素補間は画素数を疑似的に増やす画像処理技術である[27]. 補間法として一般的な方法としては最近隣補間法，線形補間法及び三次畳込補間法が挙げられる．これらの方法は出力画像の座標上に格子を設定し，その格子に対応する画素に原画像（入力画像）を対応させることで補間処理を行い，本来存在しない画素を作りだし補うものである．

最近隣補間法（Nearest neighbor）は推定する画素 y(u, v)にもっとも近い画素の信号をそのまま与える方法であり，以下の式で求められる[13].

$$y(u,\ v) = x(u + 0.5, v + 0.5)$$（式 1）[13]

最近隣補間法は原画像のデータをそのまま保存するという利点があるが，最大 1/2 画素の位置誤差が生じる[13].

線形補間法（Bilinear）は信号と信号の間を直線で結んで直線上の信号を推定し出力データとして与える方法である．二次元である画像の場合には出力画素の周辺の 4 点の原画素を用い，以下の式によって出力画素を求める[13].

$$y(u,v) = ((i+1)-u)((j+1)-v)\cdot x(i,j) + ((i+1)-u)(v-j)\cdot x(i,j+1) \\ +(u-i)((j+1)-v)\cdot x(i+1,j) + (u-i)(v-j)\cdot x(i+1,j+1)$$
（式 2）[13]

この方法は 2 の倍数以外の拡大率の場合，原画素が出力画素に保存されないという問題があるものの， 4 点の平均を求めていることから平滑化の効果もあり，アルゴリズムも簡単なことから比較的利用される処理である[13].

双三次補間法（Bicubic）は規則的な二次元空間でデータ点を補間するためのキュービック補間法の拡張である．線形補間では画素間を直線で結び，出力画素を推定するため画像全体がぼけやすい．そこで，推定する画素の周囲何点かの画像データを 3 次畳込関数を用いて求める方法が 3 次畳込補間法である． 3 次畳込関数は図 2 のような関数であり，この関数を用いることで式 3 のように出力画素を求めることができる[13].

$$y(i,j) = [f(1+q), f(q), f(1-q), f(2-q)] \begin{bmatrix} x(i-1,j-1) & x(i,j-1) & x(i+1,j+1) & x(i+2,j+1) \\ x(i-1,j) & x(i,j) & x(i+1,j) & x(i+2,j) \\ x(i-1,j+1) & x(i,j+1) & x(i+1,j+1) & x(i+2,j+1) \\ x_(i-1,j+2) & x(i,j+2) & x(i+1,j+2) & x(i+2,j+2) \end{bmatrix} \begin{bmatrix} f(1+p) \\ f(p) \\ f(1-p) \\ f(2-p) \end{bmatrix}$$

$p = (u - int(u))$
$q = (v - int(v))$

（式 3）[13]

1 次元的に輝度値のグラフ図 2.3[14]を見てみると，補間法の中で Bicubic 法は元のデータと近い特徴を再現できる方法であると考えられる．最近傍法や線形補間法よりも計算処理は重いが，画質の劣化を抑えることが出来る．

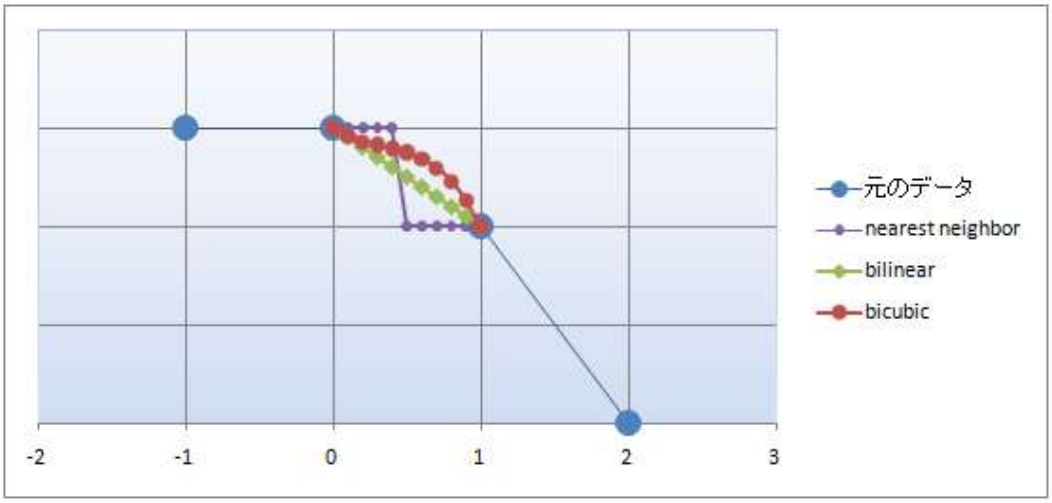


図 2.3：各画素補間法の 1 次元輝度値分布[14]
横軸は位置，縦軸は輝度値

2.2.2 SRCNN

SRCNN（Super Resolution Convolutional Neural Network）は CNN（Convolutional Neural Network）を用いて低解像度の画像と高解像度の画像の関係性を学習することにより，低解像度の画像から高解像度の画像を得るアプローチであり，従来の手法よりも高精度，かつ処理に要する時間も短いという特徴を有する[15].

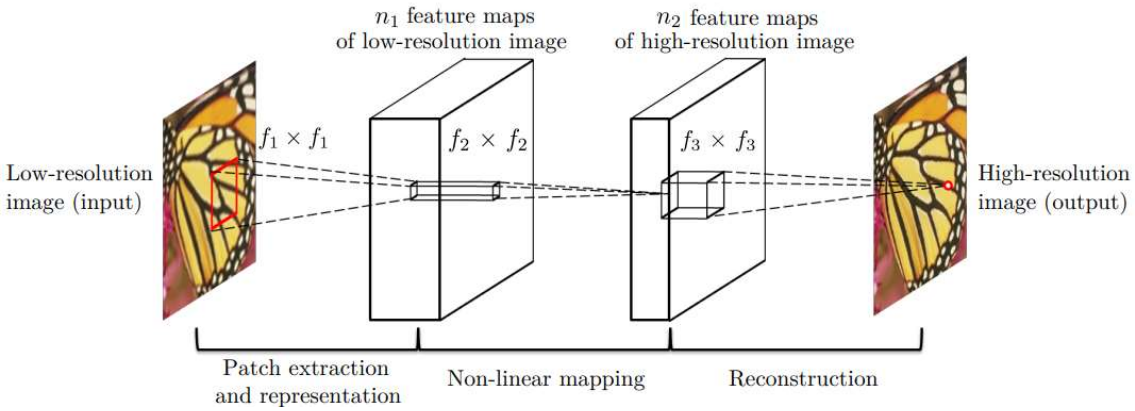


図 2.4：SRCNN モデル[16]

第一層 CNN：入力画像の特徴を取り出す．（9 x 9 x 64kernel）

第二層 CNN：第一層取り出した特徴を非線形で表現する（1 x 1 x 32kernel）.

第三層 CNN：特徴と空間近傍の予測を組み合わせて，最終的な高解像度画像を生成する（5 x 5 x 1kernel）.

SRCNN を適用した超解像処理ツールとして waifu2x[17]がある．アニメイラストやその他の写真用の画像スケーリングおよびノイズリダクションプログラムである．本研究は waifu2x を対照群のツールとして，超解像処理の結果の対比を行う.

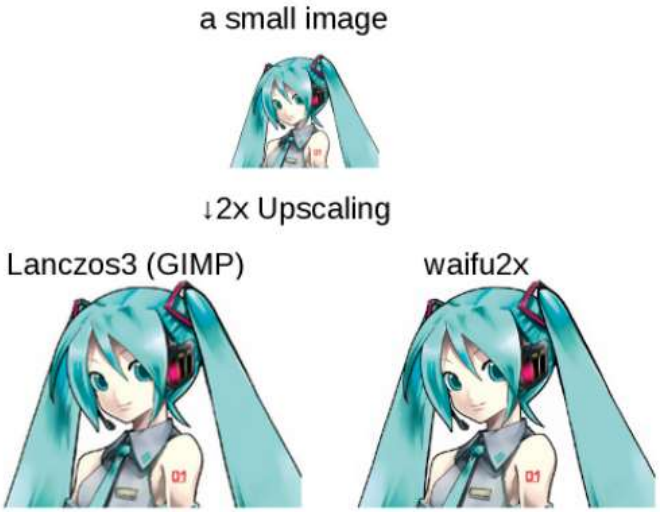


図 2.5：waifu2x の 2 倍超解像例[17]

2.2.3 SRGAN

SRGAN（Super Resolution Generative Adversarial Networks）は GAN（Generative Adversarial Networks）の手法を超解像処理領域に応用した技術であり，Ledig ら(2016)[18]によって提案された手法である．従来手法に比べて結果の知覚品質が高く，自然であるという特徴がある.

その核となる敵対的生成ネットワーク（GAN）は，Goodfellow ら(2014)の論文で, 2 つのネットワークを競わせながら学習させるアーキテクチャとして提案されたものである[11].

GAN のアーキテクチャを図 2.6 に示す．Generator は，生成データの特徴の種に相当するランダムノイズ（図 2.6 では z）を入力し，このノイズを所望のデータに近づけるようにマッピングする．Discriminator は Generator が生成した偽物のデータと本物のデータが与えられた時，その真偽を判定する[12].

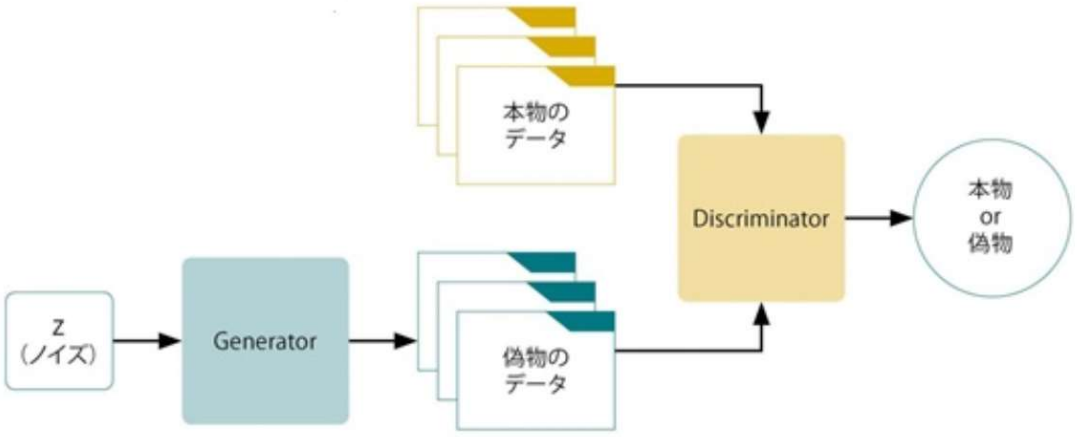


図 2.6：GAN の構造[12]

GAN の提案者である Goodfellow は「これは偽造者と警察の攻防になぞらえることができる．偽造者が本物そっくりな偽札を造ろうとするのに対し，警察は特定の紙幣を調べ，それが偽物かどうかを判別しようとするようなものである」と説明している[11]．例えば，一方が偽物の画像を見つけ出す能力を高めようとするなら，もう一方はオリジナルと見分けがつかない偽物を作成する能力を高めようとするわけである．

この 2 つのネットワークを交互に競合させ，対抗学習を進めることで，Generator は本物のデータに近い偽物データを生成できるようになる．2 つのネットワークの競合関係は，ロス（コスト）関数を共有させることで表現される．すなわち片方のロスが小さくなれば，もう一方にとってはそのロスが大きくなる．Generator はロス関数の値を小さくすることを目的に，Discriminator はロス関数の値を大きくすることを目的に学習させる[12]．この構成により，従来の生成モデルより鮮明で本物らしい画像生成が可能になった．

低解像度画像を超解像化する際に，自然に見えるパターンは図 2.7 の赤枠の画像のように，何パターンも考えられる．MSE（Mean Squared Error）ベースで考えると，どのパターンと比較して損失を計算することになったとしても，それなりに損失が小さくなるように平均的なぼやけた画像が生成されてしまう[18]．GAN では Discriminator を騙すことさえできればどんなパターンでも良いため，平均的な画像ではなく，本物っぽいシャープな画像が生成されることが期待される[19]．

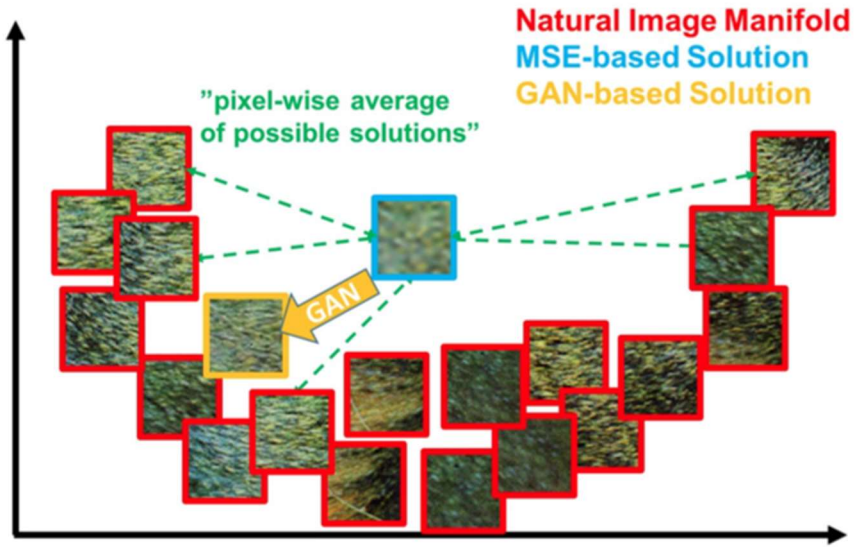


図 2.7：超解像処理のパターン[18]

Ledig ら[18]が提出した SRGAN のモデル構造は以下のような構造である．

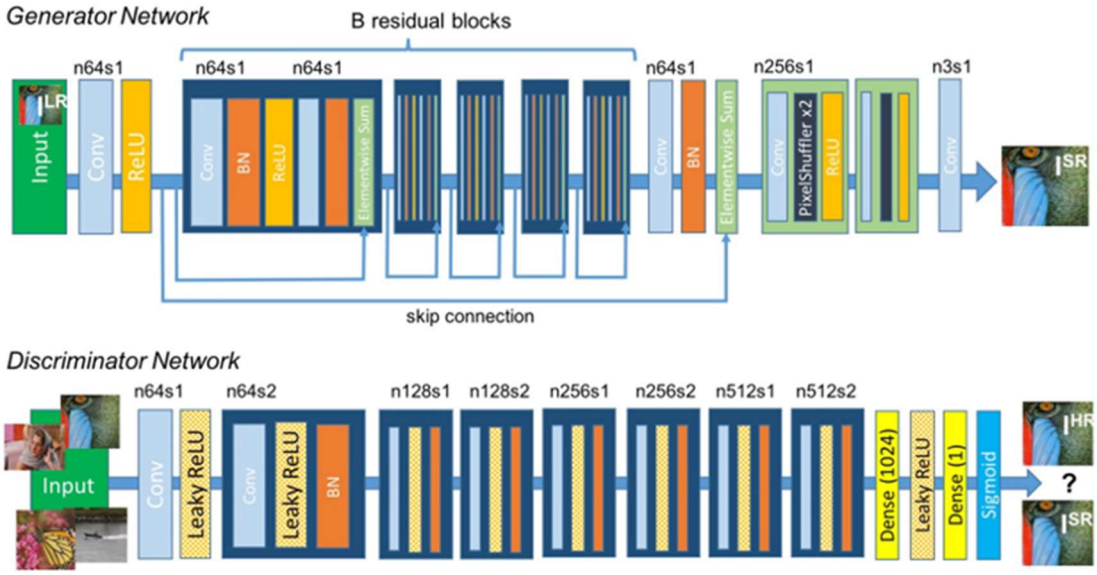


図 2.8：SRGAN の構造[18]

SRGAN モデルの生成器は ResNet を参考にしている．ResNet は，畳み込みの入出力を足し合わせて出力する残差ブロックを積み重ねたネットワークである．SRGAN でも同様な構造をしている．この構造の Skip Connection は深層学習において層の深度の限界を押し上げることができ，精度向上を果たすことが出来た．ResNet でさらに深い層を学習することが可能になった[19].

本研究はこの SRGAN モデルを基づいて，講義アーカイブシステムにふさわしいモデル構築の改善策を工夫する。さらに，モデルへの入力画像を特別な前処理手法に注目し，より講義アーカイブシステムに効果的な鮮明化板書を図る。

第3章 提案手法

3.1 全体的な流れ

本研究の流れを図 3.1 に示す.

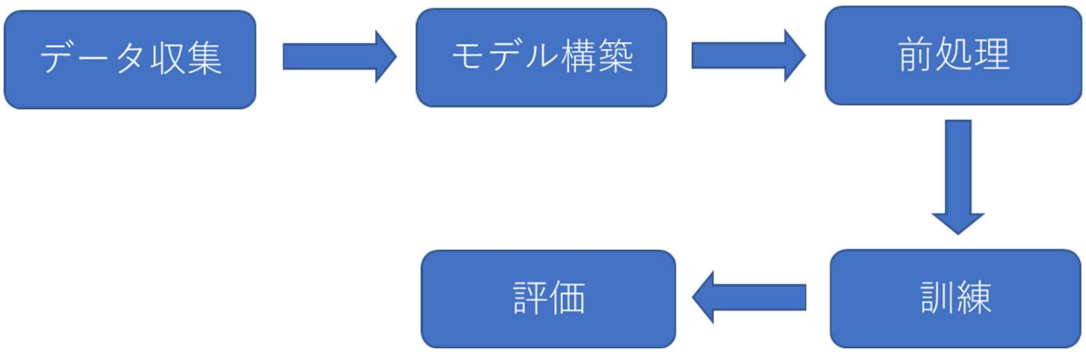


図 3.1：研究の流れ

図 3.1 の流れにより，板書の超解像処理のモデルを訓練し，実際の効果を確かめる.

3.2 データ収集

本研究は板書だけを注目することから，時間の要素を考える必要がない．そこで，全体の映像ではなく，スクリーンショットで映像内から訓練したい板書部分だけ取り出し，より効率的に訓練を行う．研究用のデータは訓練データとテストデータに分割する．ここで，鮮明化の対象が本学で配信されている HD(1,920×1,080)解像度のアーカイブのホワイトボードであるため，当該 HD アーカイブをダウンサンプリングして生成した低解像度画像から高解像度画像を生成することを目指す．講義アーカイブシステムで収録した（I239 機械学習（2018），I448 遠隔教育システム工学（2019））HD 映像の画像を入手した．テストデータは主に講義アーカイブシステムでの読みにくい画像（上記以外：I655E 現代脳計算論（2019），I470 実践的アルゴリズム理論（2018），I481 高信頼組込みシステム開発演習（2019），I217 関数プログラミング（2019））とし，提案手法と既存の手法を比較する.

3.2.1 2K 画質

本学の講義アーカイブシステムでは，HD 画質と SD(640×360)画質での配信があるが，いずれも 1,920×1,080(2K)の解像度をオリジナルとする授業映像である．そこで予備的な調査として，HD 画質映像の中で比較的ホワイトボードの内容が読みやすかった 2K 画像 142 枚を SRGAN に入力した。訓練したモデルをテストすると，以下のような結果が出た．

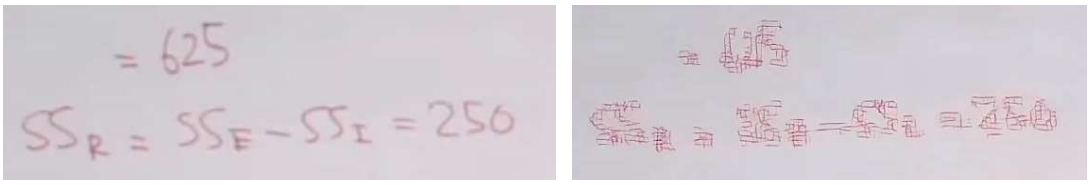


図 3.2：訓練した結果（2K）(SRGAN モデルを 2K 画像 142 枚で訓練，Epoch=10000） 左：2K テストデータ（HD），右：左の画像を 50%縮小してモデルに入力して生成した画像

右図で見られる通り，重複した線が多く生成され，読み取りにくくなっている．2K の教師データを利用する場合は，1K に縮小した画像から 2K に拡大するためのトレーニングが行われており，これを 2K から 4K へ拡大する際に適用すると，一つの線の特徴が二つに認識されてしまう等の予期しない問題が起きたものと考えられる．

3.2.2 4K 画質

本学講義アーカイブシステムでは，試験的に 4K(3,840×2,160)解像度の授業映像データが 4 つ収録されていたが，板書のバラエティが不足しており，トレーニングデータとしては不十分であると考えられた．実際の授業では，アルファベットやひらがな，カタカナ，漢字，数式や数字，マークなど様々な様式があるのみならず，異なる色のペンで重要度を示す場合も少なくない．そこで，2020 年 10 月，本学の小ホールの 4K カメラで新たな授業映像を収録することにした．今度の収録には上で挙げている様々な様式を用いた板書を収録し，より多くの場合に適用可能なモデルを得ることを目指す．

このような条件の下で予備実験として，4K 画像 53 枚を集め，SRGAN に入力した。訓練したモデルに 3.2.1 節と同じテストデータを入力すると，図 3.3 のような結果が出た．ここでは Epoch を 10000 に設定した。訓練データ量が違う

ため，2K データと同程度の結果が得られるようになるには Epoch を増やすべきだが，10000 回という学習途中の段階でも明らかに 2K のモデルより読みやすいデータが生成されたことが分かる．

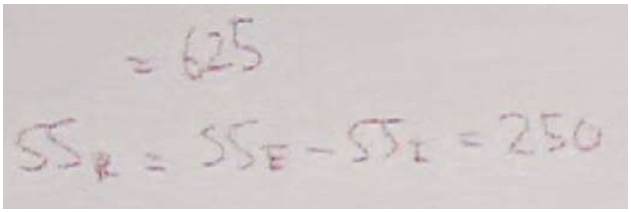


図 3.3：訓練した結果（4K）(SRGAN モデルを 4K 画像 53 枚で訓練, Epoch=10000） 3.2.1 節と同じテストデータを入力して生成した画像

最終的に，4K ビデオで収録した授業映像 10 個の中から 100 枚のホワイトボード領域のスクリーンショットを取り出した．その中で，85 枚を訓練データ，15 枚をテストデータに分割することにした．

3.3 モデル

本研究は SRGAN のモデルを踏まえ，いくつの改善策を追加した．ホワイトボードの超解像用のモデルを構築した．

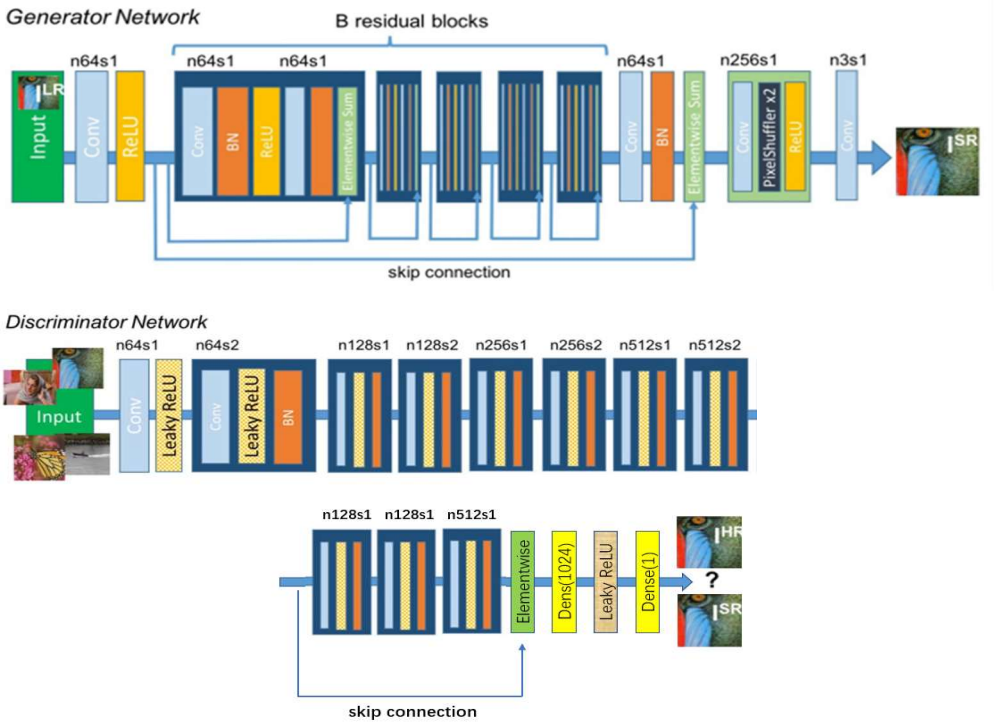


図 3.4：本研究 SRGAN 改善策モデル

モデルの中には，Generator と Discriminator がある．Generator では Conv 層，BN 層，ReLU 層，Conv 層，BN 層と Elementwise Sum 層が残差ブロックを 5 回重ねる構成となる．Skip Connection で入力の重みを出力に加え，精度を向上させる．ここまでは元の SRGAN モデルと同じである．その後は，Upsampling 層（Conv 層，Subpixel 層，ReLU 層の組み合わせ）を一個のみ残し，最後の Conv 層に処理してから画像を生成する．

Discriminator では元の 7 回ではなく，10 回の Conv 層，LReLU 層，BN 層で構成した残差ブロックの重ね合わせを中心とし，また多くなった 3 層の部分で Skip Connection を利用した．後半では Sigmoid 層を除き，2 層の Dense 層と LReLU 層だけで処理する形とした．

具体的には主に超解像サイズに合わせるモデル層の削除，過学習を防ぐため残差ブロックの増加，学習勾配消失の改善策として Activity function 層の削除などを行った．以下で本研究のモデル構成について詳しく説明する．

3.3.1 Discriminator への Skip connection 適用

Discriminator は本研究で改良したモデルを使っている．従来の SRGAN モデルでは，Skip Connection の適用は Generator のみで，Discriminator では残差ブロックしか使っていない．

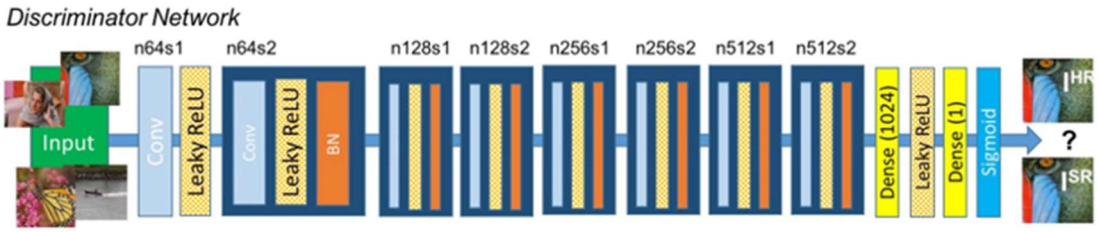


図 3.5：Ledig ら(2016)の SRGAN の Discriminator 部分[18]

zsdonghao[20]は写真に対して SRGAN の超解像処理を実行する際に，Discriminator の構造を踏まえ，ResNet の一部として Skip Connection を加えた手法を提案した．

ホワイトボードに適用する際に，どちらのモデルが適切か判断するために，二つの異なるモデルで学習実験を行った．VGG と MSE は学習過程の損失関数とし，評価対象は Generator が生成した偽りの画像と真の画像の特徴である．

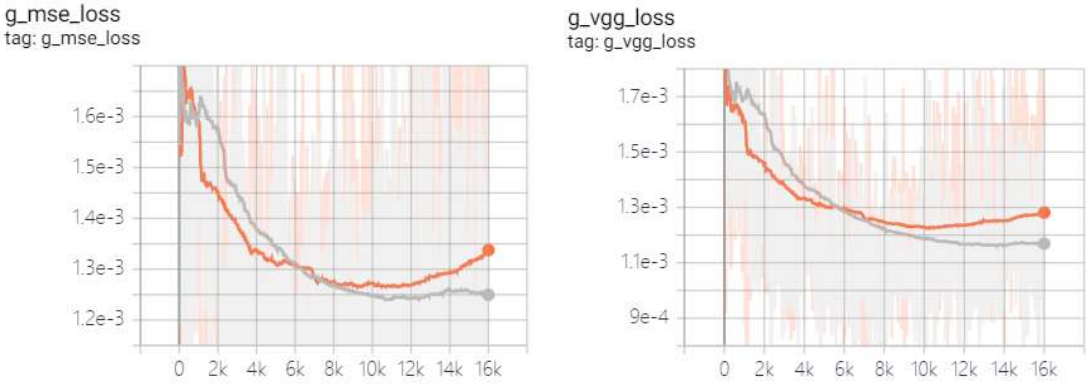


図 3.6：MSE と VGG 損失値の変化

（Epoch=16000，Smoothing=0.9）

横軸は Epoch，縦軸は損失値．オレンジ線は SRGAN モデル，グレー線は Discriminator に Skip connection を適用したモデル

図 3.6 から，SRGAN はより早く画像の特徴を掴んで学習できるが，一定の epoch 以降，loss が大きくなる過学習の兆候が見られた．Skip Connection を Discriminator に適用することは精度向上を果たす上で有用であることがわかった．

3.3.2 層の調整

SRGAN モデルは元々4 倍超解像のモデルである．実際の視聴環境では 2K 画像を拡大すればよいため，4 倍もの解像度は不要であると考えられる．訓練データは 4K データ，実際の超解像入力は 2K データであることを考えると，訓練用データを 1/2 に縮小し，2K 解像度のデータを生成して SRGAN モデルに入力することが講義アーカイブシステムに適用する上で適切だと思われる．そこで，SRGAN の Subpixel 層とそのペア，Convolution 層と ReLU 層を削除することにより，2 倍の超解像を実現した．また，出力の前の Sigmoid 層が一般画像では上手く処理に活用されたが，板書の場合はそのまま活用できないと考える．

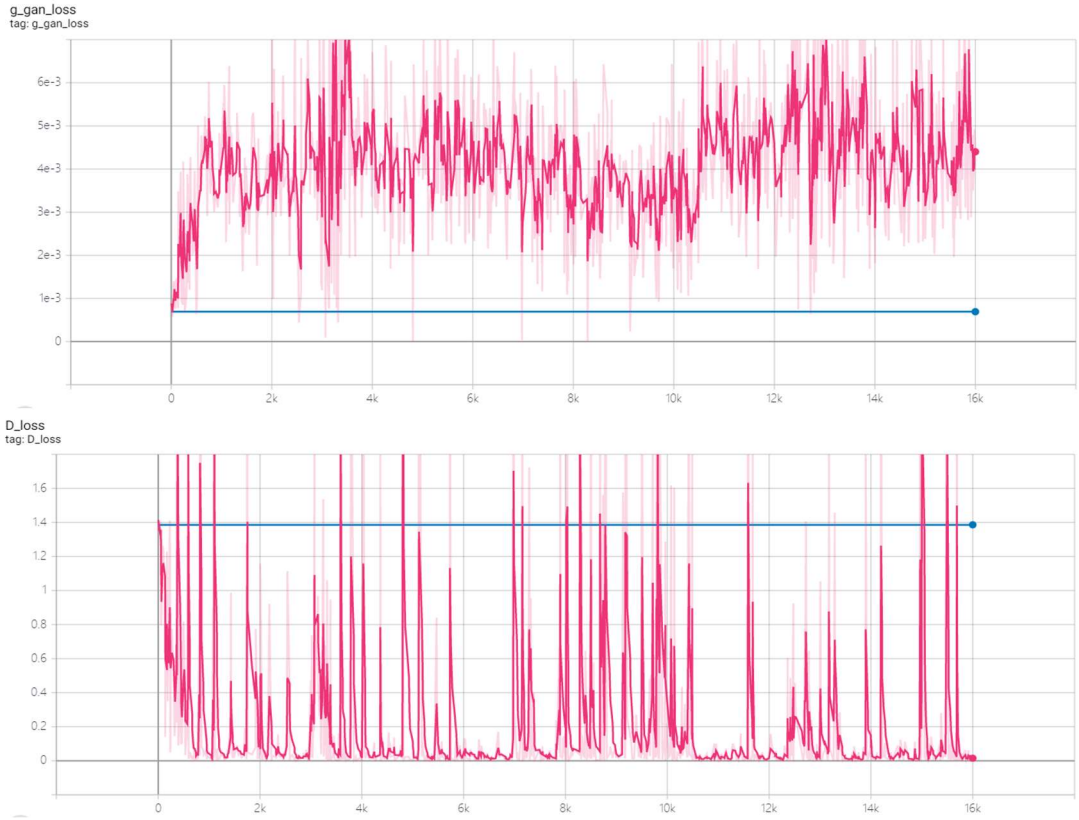


図 3.7：板書の場合 SRGAN モデルに Sigmoid 層使用の影響
青い線は Sigmoid ありの学習過程．赤い線はなしの学習過程．
横軸は Epoch，縦軸は損失値

図 3.7 は板書の場合で元の SRGAN モデルと元の SRGAN モデルの中で Sigmoid 層だけ削除した学習過程の違いを示している．g_gan loss は GAN loss であり，Generator の騙す能力を示す．値が低いことは生成した偽りの画像に良い評価を行ったという意味である．Sigmoid 層がある時，Discriminator の損失は下がらず，Generator の損失も上がらないため，お互いは対抗にはならない．このことは，Discriminator より，Generator が強すぎるため．Discriminator は生成した画像に騙され，Generator の成長につながらなかったと思われる．つまり，学習勾配消失の原因になるため，本研究では Sigmoid 層を削除した．

3.4 データの前処理

画像から必要な情報を精度良く得るために，一般的には画像の輝度やコントラスト修正，雑音除去などを行う．しかし，板書の場合には，コントラストの修正等を行うことで，板書のノイズも大きくなってしまう問題点がある．周波数空

間におけるフィルタリングは一般画像の雑音除去によく使われるが，板書の場合では周波数の差異が少なく，必要な情報とノイズはどちらでも周波数の範囲が定めにくく，フィルタリングしにくいことが容易に想像される．

入力画像はスクリーンショットで得られるため，板書領域のサイズはそれぞれ異なる．本研究では，各高解像度訓練画像を 192x192 サイズに分割することとし，その 1/2 縮小画像を低解像度訓練画像データとする．

3.4.1 正規化と反転

訓練データの特徴を最大化にする処理，いわゆる正規化処理を行う．

$$pixel\_new = \left(\frac{pixel}{127.5}\right) - 1 \quad (式\ 4)$$

処理後の画像の画素値の範囲は[－1,1]である．画素値の変化範囲を縮小することにより，ネットワークの重み更新の振れ幅が抑えられ，重みが収束しやすくなり[26]，モデルとして学習しやすくなるという利点がある．

モデルを生成する際に特定の方向に入力画像が偏る問題を避けるとともにデータを増やすことができると考え，90 度回転と反転を行う．そこで，tensorflow.image.random_flip でデータをランダムに上下反転，左右反転にさせ，numpy.rot90 でデータの 90 度回転処理を行う．

3.4.2 色空間で画像の分析

Ni[10]と小林ら[5]は板書に対して，二値化処理を行っている．二値化処理をすることで，簡単に板書での黒字の特徴を多く増やすことができる．しかし，講義アーカイブシステムでは，板書の文字色が重要な意味を持つことがあり，例えば文字色を変えることによって，注視してほしい部分・そうでない部分が表現されることがしばしばある．こうした色の情報を残すために，色の特徴の鮮明化を行うこととした．

RGB 座標系とは光、照明、色、色空間などを規定する国際標準化団体 CIE によって定められている 3 原色を R（red），G（green），B（blue）とする表色系である[28]．図 3.8 のような，板書と背景の特徴が混在し，輝度ヒストグラムでは特に鮮明化したい対象を特定できないといった問題がある．RGB 色空間では，人間の感覚に合った色彩に関する処理を行うことは難しいといえる．

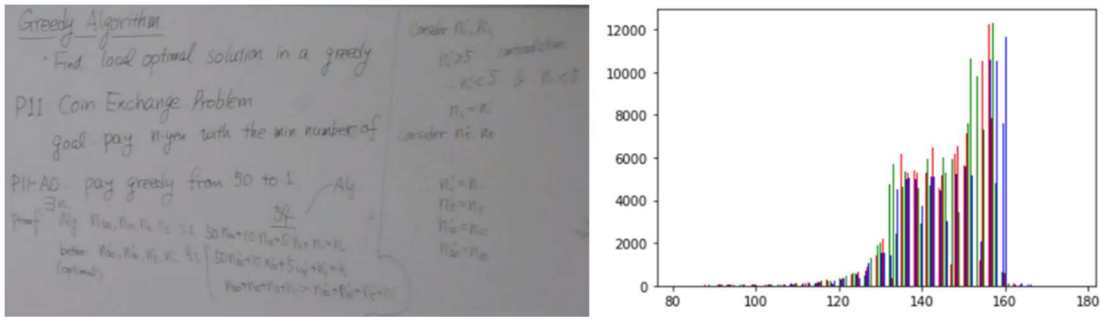


図 3.8：板書の一例とその RGB 座標系の輝度ヒストグラム
横軸は輝度値，縦軸はピクセルの数量

HSV 表色系は RGB 表色系を直観的に分かりやすいマンセル表色系に近い色相 H（hue），彩度 S（saturation），明度 V（value）に変換した表色系である[21].

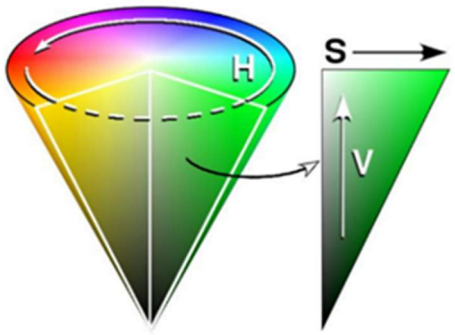


図 3.9：HSV 座標系[21]

板書と背景の HSV 値の範囲により，フィルターで処理対象を分けて処理することが可能である. HSV の値範囲は H: [0, 360] (OpenCV では[0, 180]), S:[0, 255], V:[0, 255]とした.

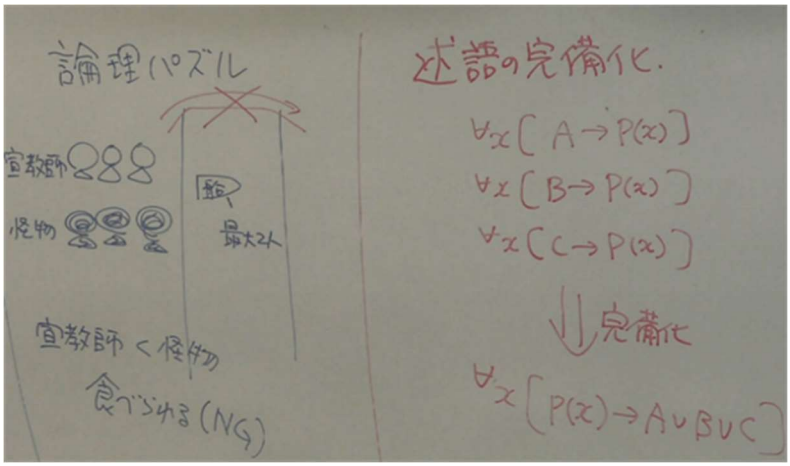


図 3.10：ホワイトボード領域の板書画像の一例
（画像おおよその HSV 値分布は以下である．赤い部分：
H:[176,10],S:[44,126],V:[122,173]；青い部分：
H:[103,120],S:[5,46],V:[90,113]；余白部分：H:[10,50],S:[5, 35],V:[104,161]）

また，図 3.10 の問題の一つは背景が暗く，コントラストが足りない点が挙げられる．そこで，10＜H＜40 のフィルターを導入し，範囲内の画素だけ輝度を上げる．

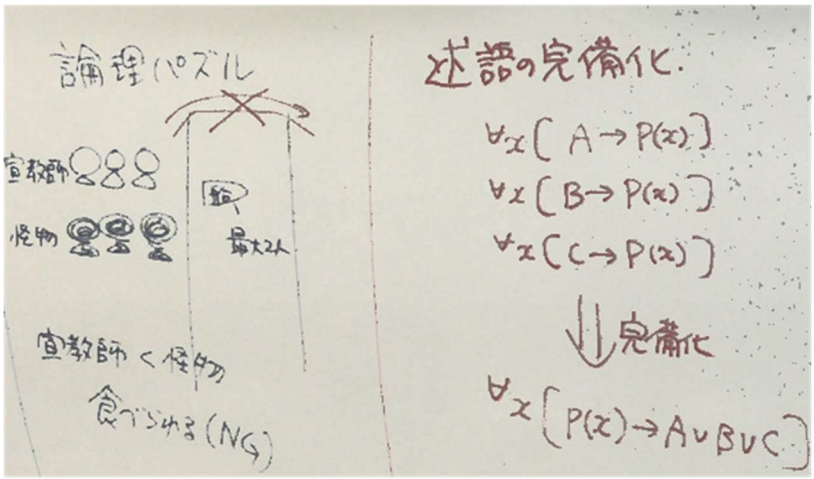


図 3.11：フィルターで処理した効果

背景色は明るくなったと思われるが，右半分に若干ノイズが残り，左半分の青い字が若干擦れた画像になっている．このように，板書の特徴を上手く強調するような前処理は大変困難であり，さらに，実際の板書環境も日光の色，教室の明るさ，文字色など様々な要因で毎回異なる．つまり，このような前処理方法は決

して汎用性が高くなく，撮影の環境，板書で使うペンの色を制限しない場合，他の前処理手法を検討する必要がある．

3.4.3 ヒストグラム均等化

ヒストグラム均等化[22]，あるいはヒストグラム平坦化は画像の強度ヒストグラムを用いてコントラストを調整する画像処理手法である．この方法は，前景と背景の両方が明るい，または暗いような画像に有効である．授業環境ではホワイトボード全体は暗い場合が多いので，前処理の一環として役に立つと考えられる．

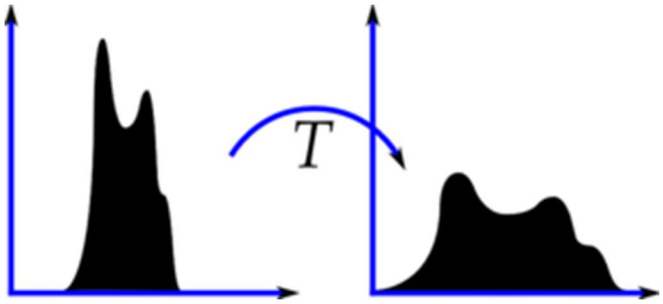


図 3.12：ヒストグラム均等化[22]

均等化手法の一つは RGB 色空間で，三つのチャンネルごとに均等化処理を行い，　最終的にそれらをマージする手法である．結果を図 3.13 に示す．

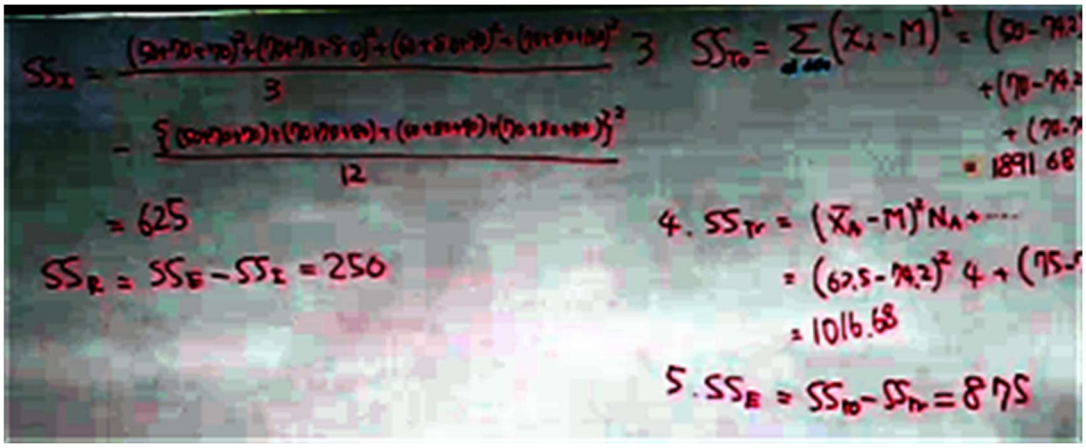


図 3.13：RGB 色空間で各チャンネル均等化処理結果

その結果，下半分のコントラストはかなり高まったが，均等化の際に，小さい画素値はもっと小さくなる傾向のため，暗いところはもっと暗くなってしまい，理想な結果であるとは言えない．さらに，各色のチャンネルごとに均等化してマ

ージすることにより，ホワイトボードの色が変わったことも確認できる．これは，各色では均等化の程度が異なるためと思われる．

そこで，CLAHE（コントラスト制限適応ヒストグラム均等化）手法[22]の適用を考える．この手法は，図 3.15 で示したように，画像を“タイル(tiles)”と呼ばれる小領域に分割し，領域ごとにヒストグラム平坦化を適用する．ノイズがある場合，ノイズも強調されるため，コントラストの制限を適用することでノイズの大幅増加を防ぐことができる．もしもビン（ヒストグラムの棒）の出現頻度が特定の上限値を超えた場合，上限値を超える画素はその他のビンに均等に分配され（図 3.14），その後にヒストグラム平坦化を適用する．平坦化の適用後にタイルの境界に生じる疑似輪郭を消すために bilinear の内挿をする[22].

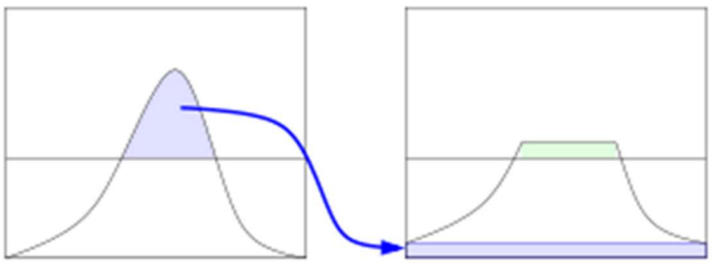


図 3.14：コントラスト制限適応ヒストグラム均等化制限値[22]

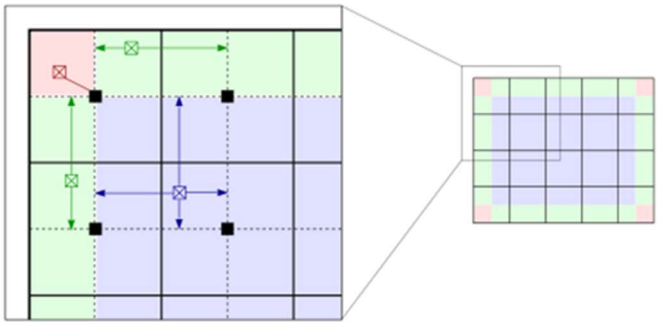


図 3.15：コントラスト制限適応ヒストグラム均等化タイル[22]

それで，OpenCV で CLAHE を適用する場合，重要な変数として clipLimit と tileGridSize の二つが挙げられる．clipLimit は制限値となり，tileGridSize は行と列のタイル数となる．図 3.16 で示した通り，clipLimit 制限値が高いほど（図 3.16 では 10.0），処理を行う制限が緩く，一般的な均等化に近づく．また，タイル数が大きいほど（図 3.16 では（50,50）），均等化処理を細かく行い，回数も多く，画素の変化が激しくなり，周波数も上昇する．

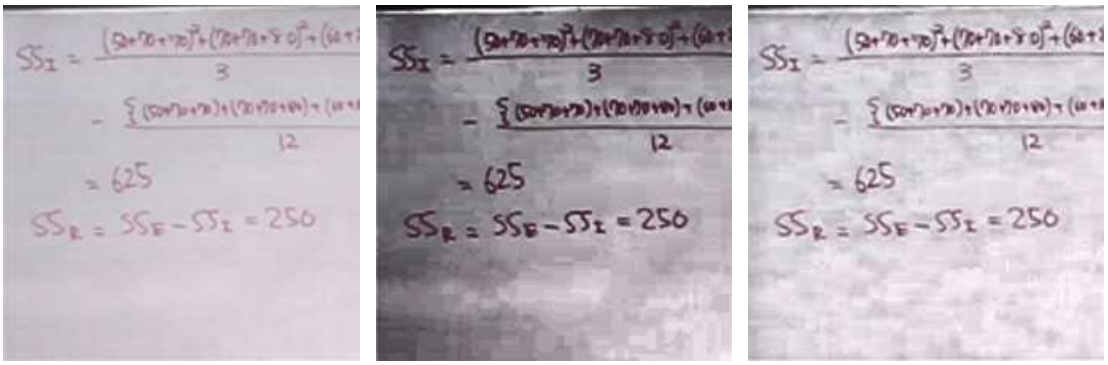


図 3.16：各 clipLimit と tileGridSize 値設定の板書効果:
1.0,(2,2);　　10.0,(2,2);　　1.0,(50,50);

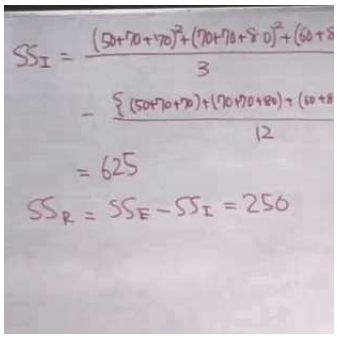


図 3.17：clipLimit=2.0,tileGridSize=(2.2)の場合

これらのことから，本研究では設定値を clipLimit=2.0,tileGridSize=(2.2)にした．図 3.17 のような，ノイズの増加をなるべく減らし，文字も鮮やかになるようにした．

3.5 訓練

学習の Epoch は事前学習 100 回と対抗学習 10000 回である. 損失関数は MSE loss, VGG loss, GAN loss で最適化する．それぞれの詳細は 3.5.3 節で示す．事前学習 100 回の訓練は画像の特徴を取り出し，MSE loss だけによって訓練する．これによって，generator はより早く学習の方向が決められ，望まない方向の過学習も防げる. 学習率を Ledig ら[18]の設定値をもとに 1e-4 と設定し, beta を tf.keras.optimizers.Adam 関数の default 値 0.9 と設定した.

Dong ら[16]の超解像処理コードでは全訓練回数の半分で学習率を 1e-1 に更新する設定だったが，GAN の不安定な要因の一つとなり，過学習がみられた．図 3.18 のように, 学習率の更新時に損失がかなり上がった. この結果をふまえ，本研究は学習率の更新を削除することにより，安定的な学習過程を確保した．

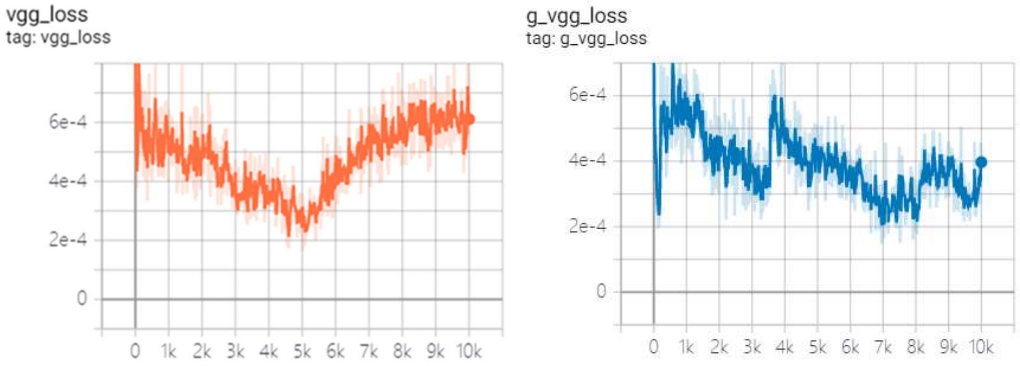


図 3.18：学習率更新により VGG 損失関数変化

横軸は Epoch，縦軸は損失値．左は学習率更新あり，右は学習率更新なし

3.5.1 SRGAN アルゴリズム

SRGAN のアルゴリズムは以下の通りである．

$$\min_{\theta_G}\max_{\theta_D} \quad \mathbb{E}_{I^{HR}\sim p_{\text{train}}(I^{HR})}\left[\log D_{\theta_D}(I^{HR})\right] + \mathbb{E}_{I^{LR}\sim p_G(I^{LR})}\left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)\right)\right] \quad \text{（式 5）[15]}$$

真の画像を 1，偽りの画像を 0 と仮定する．このアルゴリズムにより，敵対的なミニマックス問題を解決するために Discriminator は Generator とともに交互に最適化する．

Generator を訓練する時，Generator が生成した画像$G_{\theta_G}(I^{LR})$は大きければ大きいほど生成したい真の画像に近い．そして，Discriminator からの評価$D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)$が真の画像だと，1 に近い最大値を得たい．log0 は無限小であり，つまり最小値$\min_{\theta_G}$を求める．

Discriminator を訓練する時，真の画像が 1 と判断するためには，$D_{\theta_D}(I^{HR})$が最大値 1 となるように生成したい．log1 は 0 なので，$\mathbb{E}_{I^{HR}\sim p_{\text{train}}(I^{HR})}\left[\log D_{\theta_D}(I^{HR})\right]$は最大値 0 である．Generator が生成した画像を 0 と判定するためには，$D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)$は最小値としたい．$\log\left(1 - D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)\right)$からは最大値を得たい．最大値と最大値の足し算になるため，最大値$\max_{\theta_D}$を求める．

3.5.2 損失関数

本研究主で利用した損失関数は，VGG，MSE，GAN の三つである．その中で，GAN は Tensorlayer の関数 tl.cost.sigmoid_cross_entropy で計算する．

MSE と VGG は共に tl.cost.mean_squared_error で計算するが，比較の対象が異なる．MSE は Generator が生成した画像と本来の HR 画像を比較する．VGG は Generator が生成した画像の VGG19 モデルに取り出された特徴と HR 画像の特徴を比較する．

Ledig ら[18]の論文による，MSE 損失関数の計算式は以下である[18].

$$l_{MSE}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y}\right)^2 \quad \text{（式 6）}$$

VGG 損失関数の計算式は以下である[18].

$$l_{VGG/i.j}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}\left(G_{\theta_G}(I^{LR})\right)_{x,y}\right)^2 \quad \text{（式 7）}$$

GAN 損失関数の計算式は以下である[18].

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right) \quad \text{（式 8）}$$

3.5.3 Tensorboard 可視化

SRGAN は Generator で三つと Discriminator で二つの損失関数を使っている．Generator では，Discriminator を騙す能力を示す g_gan loss，生成した画像と本物画像の特徴の区別を示す VGG loss と MSE loss を利用している．Discriminator では, 真の高解像度画像を正しいと判断できる能力を示す d_real loss と，生成された高解像度画像を偽りと判断できる能力を示す d_fake loss が用いられる．多数の損失関数に訓練の時より簡易に学習の状況を把握するため，可視化を行った．ここでは，Tensorflow ライブラリの Tensorboard を使うことにした．

第4章 実験・評価

本研究は Windows10，Intel Core i9-10900K CPU 3.70GHz，NVIDIA RTX2080Ti を搭載した設備で行った．4K ビデオで収録した映像のスクリーンショット画像の中で，85 枚画像（回転後 170 枚）を訓練データとし，15 枚をテストデータにしている．Windows の Python 環境で実施し，主に使ったライブラリは Tensorlayer，Tensorflow，Opencv である．学習のスピードを向上するため, NVIDIA CUDA10.0 と Tensorflow の GPU バージョンを使った. Batch size の設定値は 8，Epoch は事前学習 100 回と対抗学習 10000 回である．損失関数の MSE loss，VGG loss，GAN loss で最適化する．

本研究訓練用のデータは 4K ビデオで収録した映像から得たため, 半分縮小した画像を用いたが，通常の講義アーカイブシステムで収録された板書のような読みにくさは再現されていない可能性がある．そこで, 最初にテストデータで評価するのではなく, 実環境での評価を実施するために, 高解像度教師データのない読みにくい板書画像を対象に評価することにした．

評価は被験者による評価実験（全体と細部），画質評価関数での客観評価とノンパラメトリック検定で行う．検定は IBM　SPSS で行う．

4.1 全体評価

鮮明化した板書全体に対する主観的な印象を評価する実験を行った．具体的には, 4 つの手法で生成された板書の拡大画像を読みやすい順で並べる問題である．全 8 問で，各問題では同じ入力画像に対して 4 つの手法：本研究 SRGAN，元 SRGAN，SRCNN，BICUBIC による出力を見せ，読みやすい順に並び替えをさせた．実際の講義アーカイブシステムの視聴者である，本学学生 13 名からデータを収集した. 評価環境を統一するために, 被験者が評価に参加した部屋は日光の影響に及ぼしにくく，画面の設定も一定とした．また，入力の画像の読みにくい原因は様々なものがあるが, 主に字と背景のコントラストが低い，全体的なぼやけ，字が薄い，線が細い，字が小さいなどを含める．さらに，一つの画像は必ず一つの独立した原因と対応するとは限らず，複数の読みにくさの原因を同時に含む場合が多いが，分析の際に重要な原因を取り出して考慮する．

4 つの結果を受験者に個人の主観視点で読みやすい順で並べさせた結果を度数分布としてまとめたものを図 4.1 に示す．各手法の読みやすさの順位に対して，SPSS により Friedman 検定を行った結果，$\chi 2=53.9$ (p=0.00)であり，全ての条件で差がないという帰無仮説は棄却された．そこで，多重比較を行い Bonferroni の方法で調整した結果，中央値は，提案手法＞SRCNN≒SRGAN>BICUBIC の差があることがわかった．

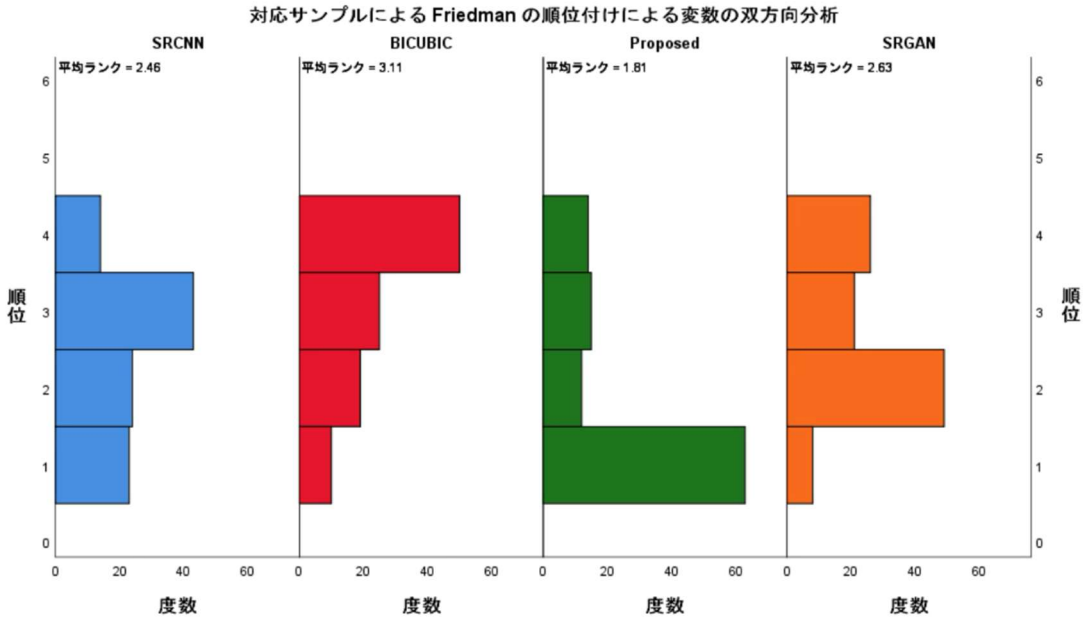


図 4.1：SPSS による Friedman の順位の度数分布

表 4.1：各手法のペアごとの比較

ペアごとの比較

Sample 1-Sample 2	検定統計量	標準誤差	標準化検定統計量	有意確率	調整済み有意確率[a]
Proposed-SRCNN	.654	.179	3.652	.000	.002
Proposed-SRGAN	-.817	.179	-4.565	.000	.000
Proposed-BICUBIC	1.298	.179	7.251	.000	.000
SRCNN-SRGAN	-.163	.179	-.913	.361	1.000
SRCNN-BICUBIC	-.644	.179	-3.598	.000	.002
SRGAN-BICUBIC	.481	.179	2.685	.007	.043

各行は、サンプル 1 とサンプル 2 の分布が同じであるという帰無仮説を検定します。
漸近的な有意確率 (両側検定) が表示されます。有意水準は .05 です。

a. Bonferroni 訂正により、複数のテストに対して、有意確率の値が調整されました。

これらの結果から，本研究提案した改善策を適用した SRGAN（Proposed）は講義アーカイブシステムにおける超解像度処理で被験者から最も良い印象を得た．ただし，詳細に結果を分析すると課題も発見された．

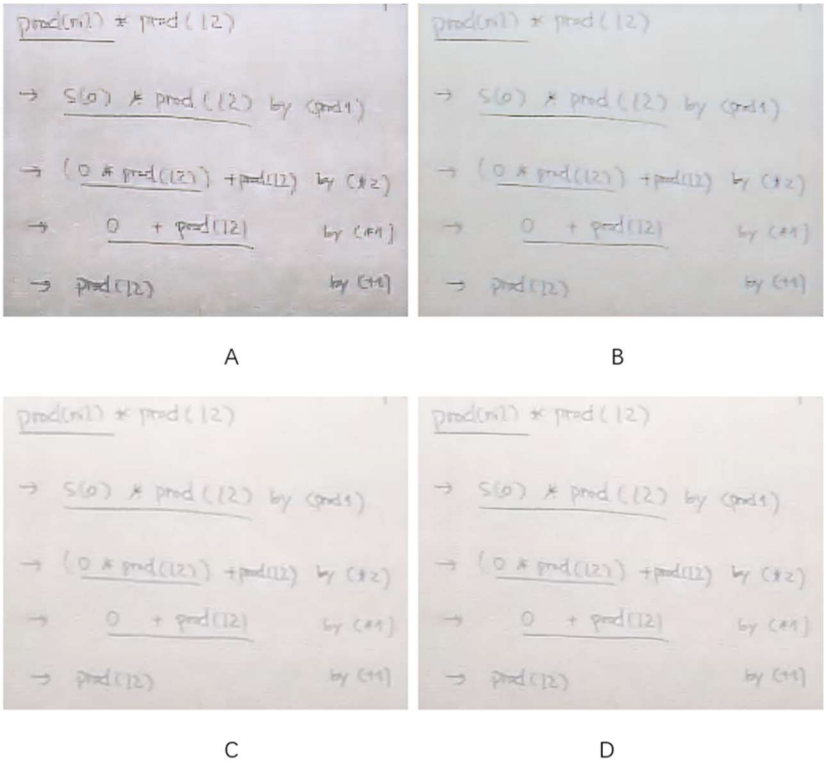


図 4.2：考察できる全体の評価問題の一例
問題の選択肢 A:Proposed，B:SRGAN，C:SRCNN，D:BICUBIC

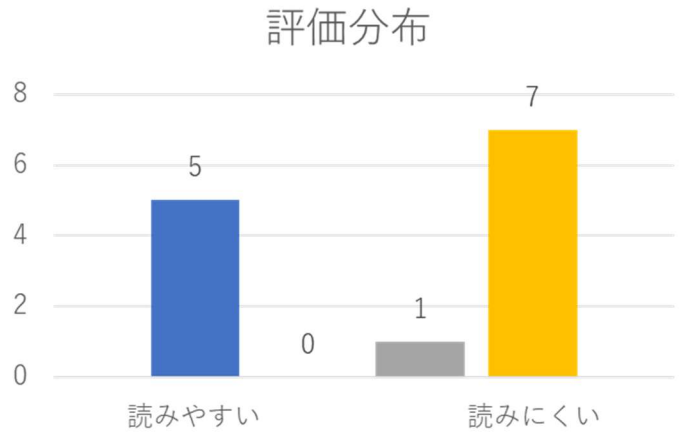


図 4.3：A（提案手法）に対しての 4 段階評価分布
横軸は読みやすさ，縦軸は人数

図 4.3 において，A(提案手法)を一番読みやすいと判断した被験者が 5 名であったのに対して，一番読みにくいと判断した被験者は 7 名であった．このように，両極端な評価となった画像があることが確認できた．これは，ノイズに対して人によって感覚が異なり，評価の基準が違うためであると考えられる．一般に，線の色が濃くなるとともに，ノイズも増加し，背景も汚くなる．とりあえずきれいな板書が欲しい人と，ぼやけない，コントラストが第一と考える人がいることが両極分布の評価に示されている．

4.2 細部評価

4.1 節の全体評価で全体的な読みやすさの程度を示し，特に全体的なぼやけや，コントラスト低下，線が薄いなどの場合においては，高く評価された．しかしながら，全体的に見る場合，読みにくいところの前後や，専門用語，言語の文法に影響される可能性が高く，単一の文字の再現度について同様な効果が示せるとは限らない．そこで，板書の細部の再現度に注目し，特に文字が小さく，特徴量が少ない場面を評価する．

この実験では，提案手法と 4.1 節で 2 番目の評価が得られた SRCNN 法で生成された板書の拡大画像から細部の文字を取り出して，4.1 節の実験に参加した被験者に読み取らせる実験を行った．両手法について，各 10 枚の同じ内容の極端に読みにくい板書画像を設定した．なお，被験者が内容を推測できないように，全ての画像を被験者毎にランダムな順序で提示した．

この実験の結果は正解率で計算する．本研究の提案手法による正解率は 31%，SRCNN の正解率の平均値は 39%であった．

13 人の正解率に対して Friedman 検定で行った結果，$\chi 2=4.5$ (p=0.034)であり，5%有意で分布が同じであることが棄却された．つまり，SRCNN の方が部分的に切り出した場合は細部を正しく読むことができる被験者が有意に多かったと言える．

また，図 4.4 に正解数の人数分布を示す．提案手法と SRCNN はいずれも 20%(2 問)正解の被験者が多かった．これらの結果から，いずれの手法であっても特徴量が少ない小さな文字に対しては，超解像処理の効果が限定的であることが確認できる．ただし，板書を書いた教員の筆跡の癖などにより認識しにくい可能性があることにも留意する必要がある．二番目の正解率について，本提案手法は 30%，SRCNN は 70%であり，この差が手法の差であると考えられる．4.1

節に述べた通り，SRCNN は SRGAN より画像の再現性が強く，文字の特徴量が少ない場合そのままぼやけて出力されるケースが多い．SRGAN はその可能なパターンの一つだけ出力するため，シャープな画像が生成されるが，特徴が変化することがある．その結果がこれらの差となって表れたと考えることができる．

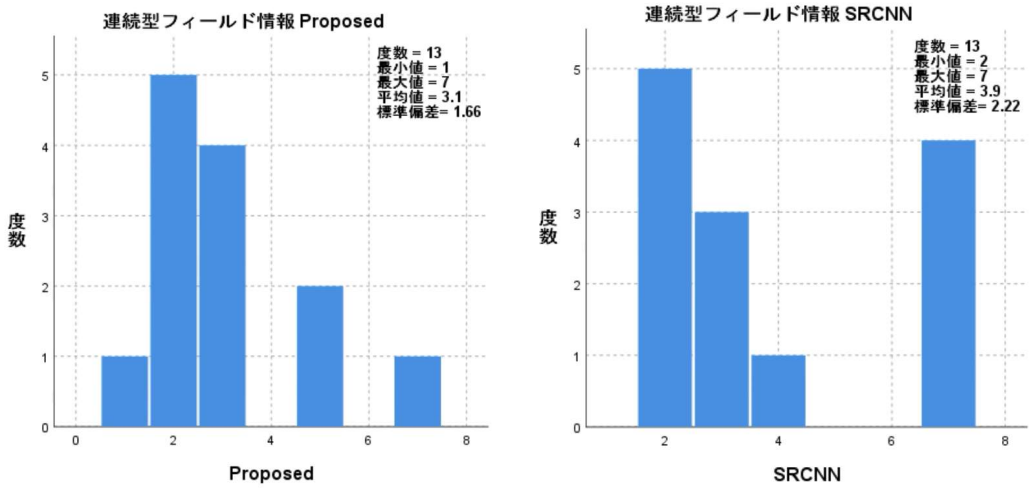


図 4.4：提案手法と SRCNN の細部読取課題の正解数の人数分布
横軸は正解率，縦軸は人数.

もちろん，板書の文字を大きく書くと，文字の特徴量が多くなる．SRGAN により特徴がある程度変わっても，認識に必要な特徴量が十分に残っていれば正しく認識できる可能性がある．文字がそこまで小さくない板書では別の文字に認識されるケースは限られると考えられる．その場合は，提案手法と SRCNN の正解率の差はさらに少なくなると予想できる．

4.3 PSNR 画質評価

4.2 節の細部の文字の再現結果は良好でなかったため，追加の各手法の再現性考察として，PSNR 画質評価も行った．

ピーク信号対雑音比 PSNR（Peak Signal-to-Noise Ratio）は画質評価関数としてよく画像処理領域で使われている．その意味は変換後の画像がどの程度劣化したかを客観的に評価する指標となる．計算式は以下である．

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right)$$（式 9）

単位はデシベル（dB), MAX は元画像がとりうる最大画素値のことである[23]. また, この式の中であらわされる MSE とは平均二乗誤差（Mean Square Error）のことであり，以下の式になる．

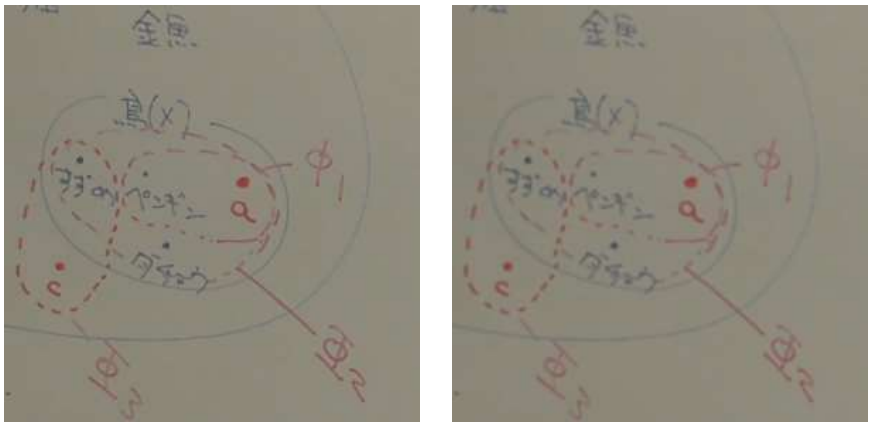
$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|X(i,j) - X'(i,j)\|^2$$（式 10）

m,n が画像の縦，横のサイズ．X が元画像，X'が劣化画像を示す[23].

ここで，手法とモデル自体だけを注目したい．本研究で利用した CLAHE 前処理手法は画質の劣化原因の一つとなるため, 画質評価では, テストデータを入力する前に CLAHE 処理を行わないこととした．

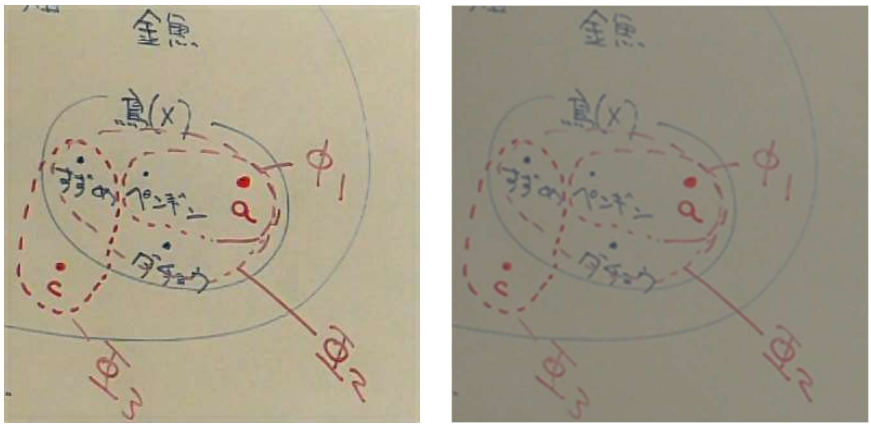
それで，本研究の 15 枚テストデータを各手法に入力した．本研究 SRGAN 改善手法と SRCNN，BICUBIC，3 種類のモデルで生成した結果の一つを図 4.5 に示した．

SRCNN　　BICUBIC



(39.29dB)　　(38.75dB)

本研究 SRGAN　　Original



(16.88dB)

図 4.5：原本と各モデルの超解像処理出力結果

左からは waifu2x の SRCNN，BICUBIC，本研究 SRGAN，原本.

表 4.2 に 15 枚のテストデータに対する，SRGAN と SRCNN，BICUBIC による拡大処理時の PSNR の平均値と標準偏差を示す. BICUBIC は平均値 27.86, 標準偏差 3.99;SRCNN は平均値 27.29, 標準偏差 4.46;SRGAN は平均値 19.10, 標準偏差 1.98 であり、手法の再現性は BICUBIC≒SRCNN＞SRGAN ことが分かった.

2.2.3 節で述べたように，MSE 損失関数ベースの手法では，どのパターンと比較して損失を計算することになったとしても，それなりに損失が小さくなるように平均的なぼやけた画像が生成される特徴がある[18]. SRGAN は対抗学習の損失関数ベースで, そのパターンの中で一つだけ生成することにより，シャー

プな画像が生成できるが，細部の特徴を間違えて生成した可能性があることを改めて確認した.

表 4.2：異なる手法によりテストデータ 15 枚の PSNR 値，平均値（AVG）及び標準偏差（STDEV.P）

	bicubic	srcnn	srgan
1	28.132	26.896	20.627
2	22.98	22.302	19.02
3	23.847	23.001	19.436
4	24.273	23.473	18.41
5	26.736	26.022	18.626
6	23.874	22.613	16.492
7	24.707	23.461	18.308
8	31.443	30.805	22.841
9	31.565	31.752	23.742
10	29.288	29.305	17.009
11	29.958	30.661	19.176
12	26.749	26.154	18.377
13	29.427	28.928	19.942
14	26.167	24.753	17.614
15	38.755	39.293	16.883
AVG	27.86007	27.2946	19.1002
STDEV.P	3.985096	4.459551	1.980859

（注：srgan は CLAHE 前処理適用なしの本研究提案手法）

第5章 おわりに

本研究では高等教育機関で導入されている講義アーカイブシステムにおいて，ホワイトボードの板書が読みにくい問題に着目した．従来の研究で提案されてきたハードウェアによる解決方法では，新たな機器を全ての部屋に導入するには時間とコストがかかることや，4K などの高精細な動画を配信するとネットワーク帯域が非常に大きくなるため，現実的ではない．また，既存のシステム内の講義動画も鮮明化できない．JAIST では講義アーカイブシステムを 2006 年から導入しており，収録した映像がすでに多くあるため，ソフトウェア上での鮮明化手法がより有意義だと考えられる．

また，講義アーカイブシステムに適用可能であると考えられる超解像処理ソフトウェア手法がいくつある．近年 SRCNN，SRGAN に代表される深層学習の技術が盛んで，一般的な画像の超解像処理手法として十分な効果を示した．ただし，板書の場合でも同じような効果が得られるとは限らない．さらに，講義の特質としては，板書の変化が激しくなく，映画やアニメと比較して書かれた文字が残っている時間が長い点が挙げられる．このため，超解像処理の対象として，動画から板書領域を取り出したスクリーンショットを対象とした．本研究では，よりシャープな画像が生成でき，超解像処理を行った際の高い鮮明化効果を期待して，SRGAN による処理を適用した．

また，本研究では，SRGAN を講義アーカイブに適用するにあたって，いくつかの改善策を適用した．モデルとしては，2 倍超解像処理を行うため，Subpixel 層と Convolution 層と ReLU 層各一層，また学習勾配消失の原因として Sigmoid 層を削除した．さらに，精度向上するために，Discriminator に Skip Connection を追加した．一般的な前処理を行う上で板書データ特有の前処理，コントラスト制限適応ヒストグラム均等化も使用し，効果的なパラメータを発見した．このような調整により，SRGAN 手法よりノイズが多少上がるが，コントラストと色の濃さが大幅に増加することができた．評価実験で SRCNN，BICUBIC などの手法を超え，被験者による読みやすさの評価が最も良い結果となり，提案した改善策の妥当性が示された．

しかし，GAN 系の超解像処理はシャープな画像を生成することに伴い，新たな望ましくない特徴を生成することがある．このことにより，小さな文字の細部の読み取りに支障を与える．提案手法と SRCNN を比較したところ，CNN 系の

超解像処理手法の方が再現性が高く，そのまま細部の特徴を保留することができる．ただし，本研究の細部客観評価から，いずれの超解像手法でも 40%未満の正解率であり，特徴量が少ない文字の読み取りに対しては，より良い超解像処理手法を検討する必要がある．

今後の課題としては，まず SRGAN で高精細化した板書における再現率を向上させることである．本研究提案手法は SRGAN 手法を核として板書を読みやすく処理したが，細部については，SRCNN の再現性の高さという優れた点を参考にできるではないかと考える．また，4 章では良い効果が得られた CLAHE 前処理は画質の劣化に繋がり，ノイズが生成され，ノイズを気にする被験者には適さないことが確認された．そこで，ノイズの消去をセットにした手法を検討することが一つの方向性として考えられる．最後に，SRGAN で使われている損失関数の中で VGG と MSE がある．この生成した画像と元の高解像度画像の特徴対比で生成した損失はホワイトボード領域の余白部分に影響され，学習の深度を深まらないという課題が見られた．このことから，板書に合う損失関数の設定も必要であると考えられる．

謝辞

本研究を進めるにあたりご指導を頂いた主指導教員の長谷川忍准教授，研究室のミーティングで多くの助言をいただいた太田光一助教に心より感謝致します. また, 貴重な意見を下さった池田心准教授, 岡田将吾准教授に感謝致します. 最後に，心の支えとなった友人の皆様に感謝致します.

研究業績

口頭発表

1. CUI BOWEN，太田　光一，長谷川　忍
   敵対的生成ネットワークによる講義アーカイブシステムの板書鮮明化, 人工知能学会先進的学習科学と工学研究会，2021 in press.

参考文献

[1] 吉良元, 長谷川忍, 大学院生の補完的学習環境としての講義アーカイブシステムの運用と分析, 教育システム情報学会誌, vol.32, no.1, pp.98-110(2015)

[2] 2017 年度 1-1 期アンケート結果-北陸先端科学技術大学院大学 情報社会基盤研究センター, 遠隔教育ユニット http://dlc.jaist.ac.jp/enkaku/htdocs/?page_id=313

[3] 長谷川忍, 2.1 遠隔学習システム, 教育工学選書 II 第 1 巻 e ラーニング／e テスティング, ミネルヴァ書房, pp.33-48(2016).

[4] 長谷川忍, 辻誠樹, 但馬陽一, 宮下和子, リモート管理・自動運用を志向した講義アーカイブシステムの開発と運用, 教育システム情報学会研究報告, vol.28, no.7, pp.49-54(2014)

[5] 小林弘彬, 長谷川忍, 講義アーカイブシステムにおけるホワイトボード領域の鮮明化. 先進的学習科学と工学研究会, vol.78, pp.30-33(2016)

[6] 大西正輝, 村上昌史, 福永邦雄, 状況理解と映像評価に基づく講義の知的自動撮影, 電子情報通信学会論文誌, D-II, vol.J85-D-II, no.4, pp.594-603(2002)

[7] 市村哲, 福井登志也, 井上亮文, 松下温, Web 学習用コンテンツを自動作成する板書講義収録システム, 情報処理学会誌, vol.47, no.10, pp.2938-2946(2006)

[8] 矢田裕紀, 鶴岡信治, 吉川大弘, 篠木剛, 遠隔授業映像撮影のためのカメラ映像と板書画像を併用したカメラ視野の決定法, 電子情報通信学会技術研究報告, 電子情報通信学会. vol.103, no.585, pp.89-94(2004)

[9] 西原功, 島田敏一, 中野慎夫, タブレット端末による簡便な課外遠隔講義システムの構築, 電子情報通信学会技術研究報告, 電子情報通信学会. vol.113, no.482, pp.89-94(2014)

[10] Wode Ni, Whiteboard Scanning Using Super-Resolution. Dickinson College Honors Theses. pp.221(2016) http://scholar.dickinson.edu/student_honors/221

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Nets.(2014)

https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[12] 平内雅則，GAN：敵対的生成ネットワークとは何か～「教師なし学習」による画像生成．IS magazine，no.20, (2018)

[13] 画像の拡大法
http://www.f-kmr.com/PDF/dsp_interpolate.pdf

[14] 画素の補間の計算方法
https://imagingsolution.net/imaging/interpolation/

[15] 川嶋一誠，深層学習を用いた衛星画像の超解像手法，Journal of The Remote Sensing Society of Japan．vol.38, no.2, pp.131-136(2018)

[16] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol.38, no.2, pp.295-307(2015)
https://arxiv.org/abs/1501.00092

[17] waifu2x
https://github.com/nagadomi/waifu2x

[18] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.4681-4690(2017)

[19] Shinya yuki，はじめての GAN，Elix Tech Blog
https://elix-tech.github.io/ja/2017/02/06/gan.html

[20] zsdonghao. Super Resolution Examples.
https://github.com/tensorlayer/SRGAN

[21] WIKI HSV 色空間
https://ja.wikipedia.org/wiki/HSV%E8%89%B2%E7%A9%BA%E9%96%93

[22] OpenCV-Python Tutorials 1 documentation ヒストグラム その 2: ヒストグラム平坦化
http://labs.eecs.tottori-u.ac.jp/sd/Member/oyamada/OpenCV/html/py_tutorials/py_imgproc/py_histograms/py_histogram_equalization/py_histogram_equalization．html#histogram-equalization

[23] 画質指標 PSNR を求める．http://who7s.blog.shinobi.jp/

[24] 保髙隆之, 山本佳則, ユーザーからみた新しい放送・通信サービス:2018 年 11 月メディア利用動向調査の結果から, NHK 放送文化研究所, 放送研究と調査, vol.69, no.7, pp.36-63(2019)

[25] WIKI e ラーニング
https://ja.wikipedia.org/wiki/E%E3%83%A9%E3%83%BC%E3%83%8B%E3%83%B3%E3%82%B0

[26] 正規化 標準化(機械学習)の理由, 必要性, メリットと元に戻す(逆変換)方法を Python で解説.
https://nine-num-98.blogspot.com/2019/12/ai-normalization.html

[27] 画素補間. goo 辞書
https://dictionary.goo.ne.jp/word/%E7%94%BB%E7%B4%A0%E8%A3%9C%E9%96%93/

[28] 田中敏幸, 画像情報処理の基礎, コロナ社(2019)

付録

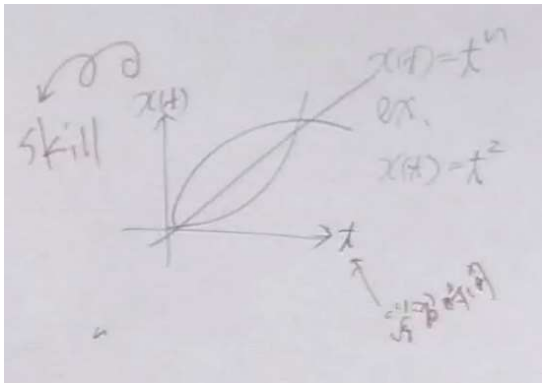
主観評価

以下の画像を読みやすい順、または実際視聴したい順で並べて下さい。Please sort the following photos from (Easy to read) to (Hard to read).

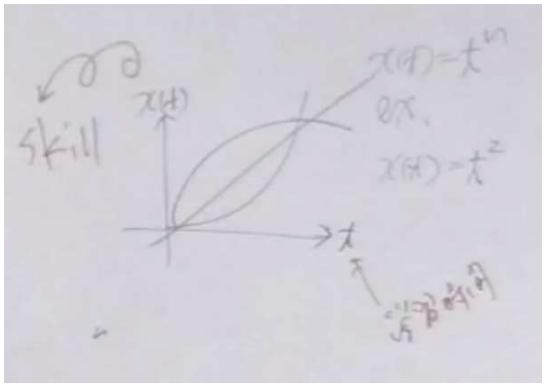
例：読みやすい（ A， B ， C ，D ）読みにくい

Ex. Easy (A, B, C, D) Hard

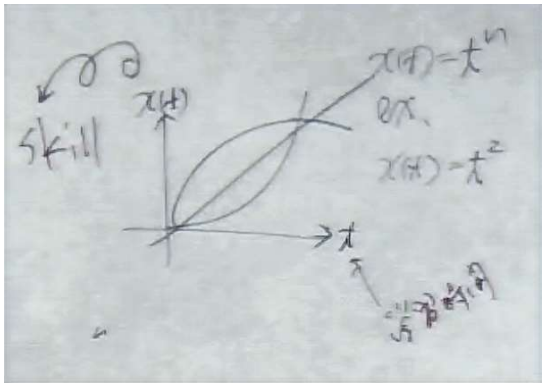
1. 読みやすい（ ， ， ， ）読みにくい



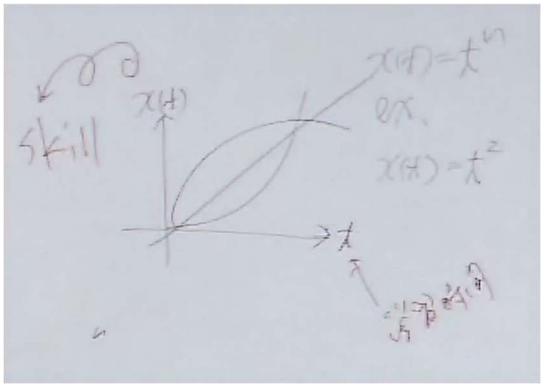
A



B

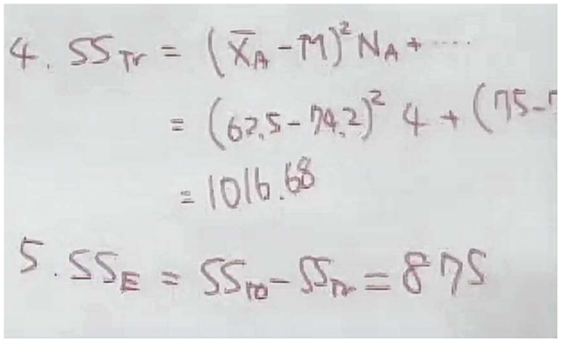


C

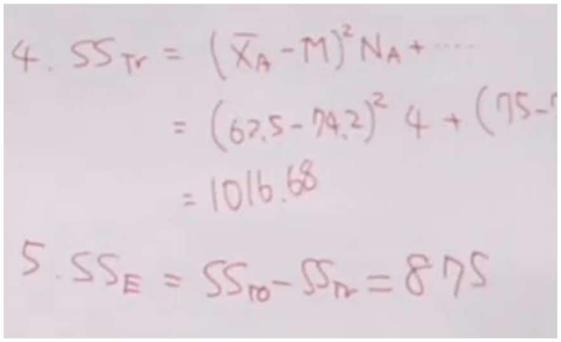


D

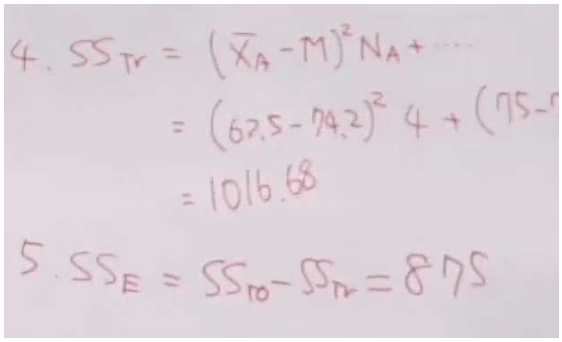
2. 読みやすい（　　,　　,　　,　　）読みにくい



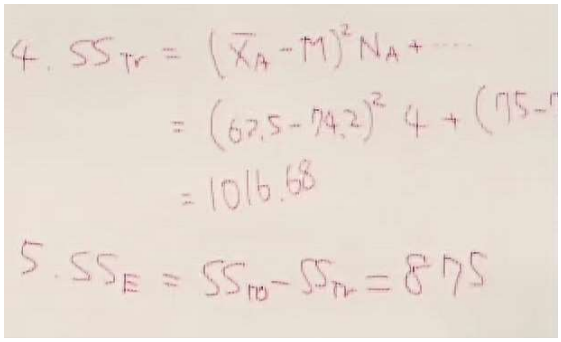
A



B

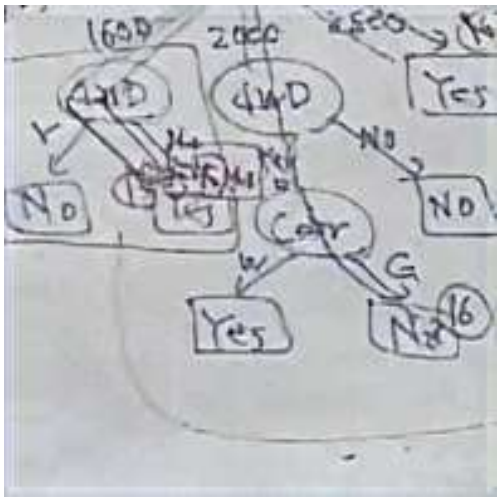


C



D

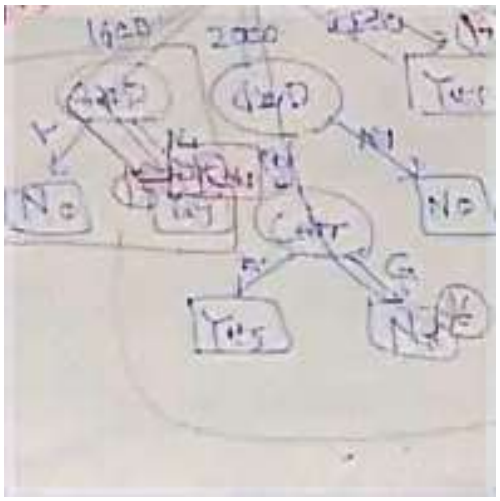
3. 読みやすい（　　,　　,　　,　　）読みにくい



A

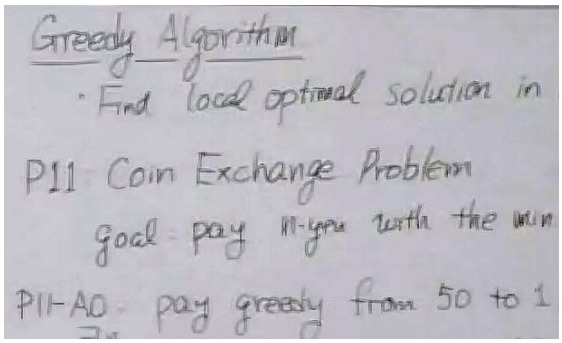
B

C

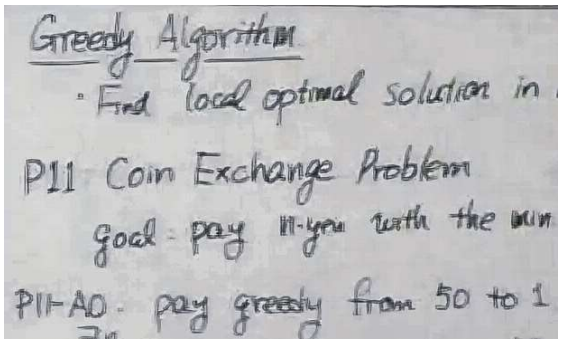


D

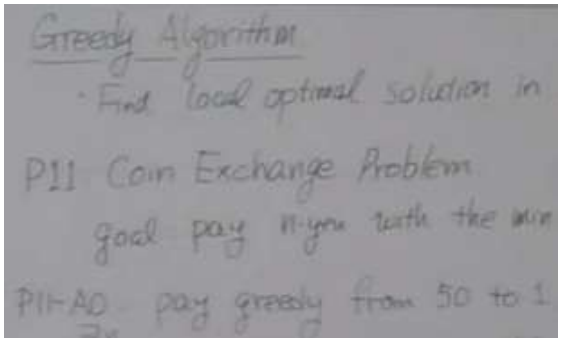
4. 読みやすい（　　,　　,　　,　　）読みにくい



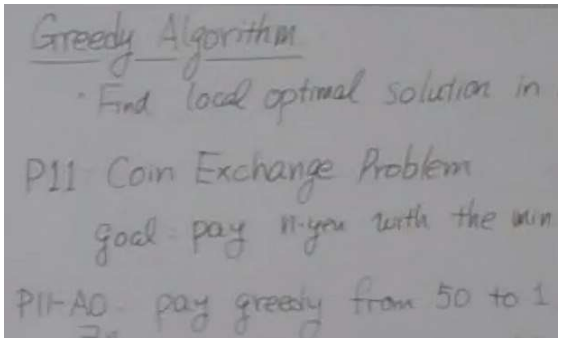
A



B

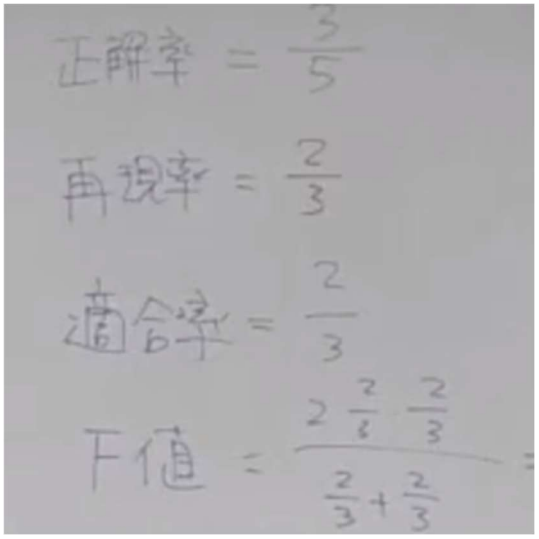


C

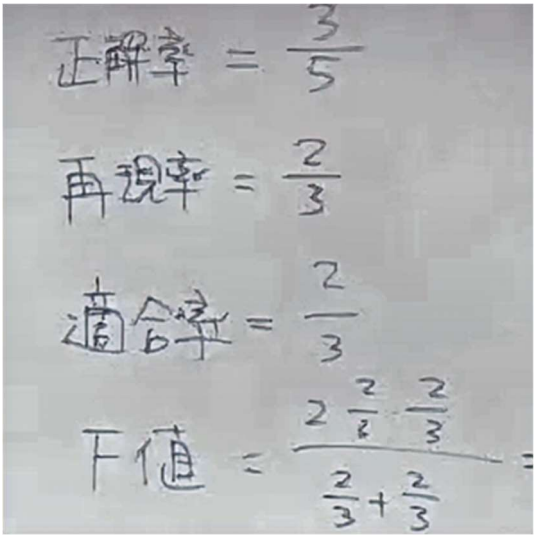


D

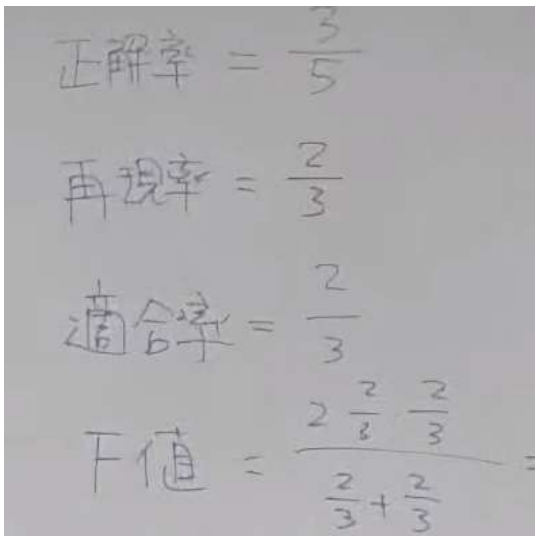
5. 読みやすい（　　,　　,　　,　　）読みにくい



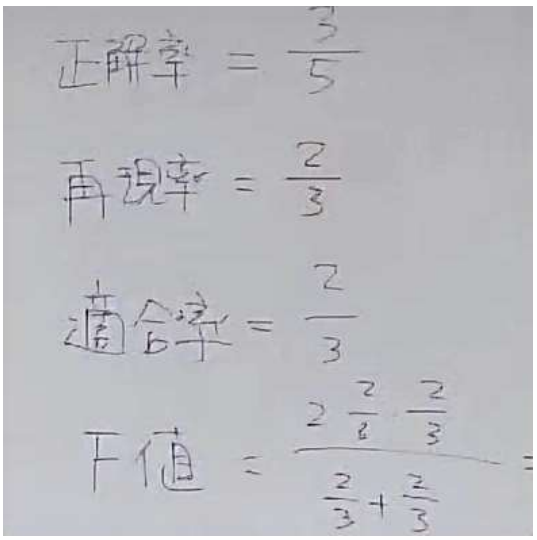
A



B

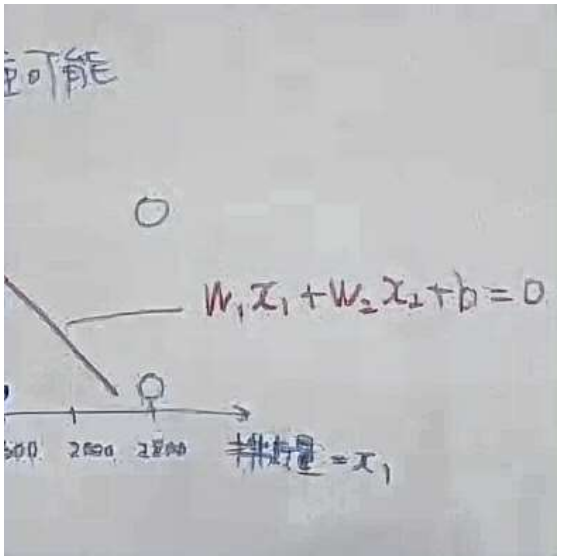


C

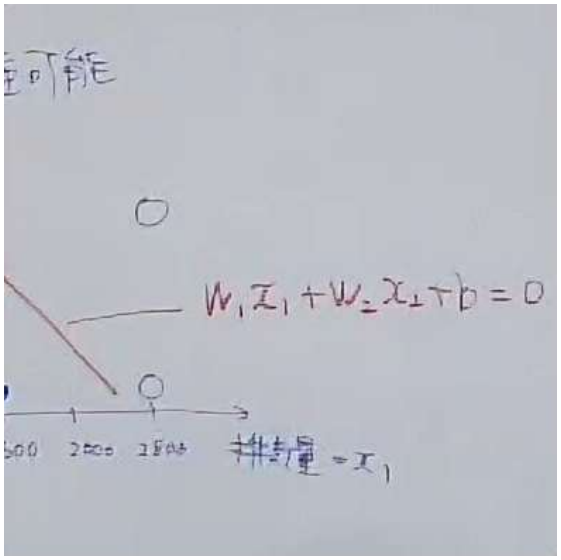


D

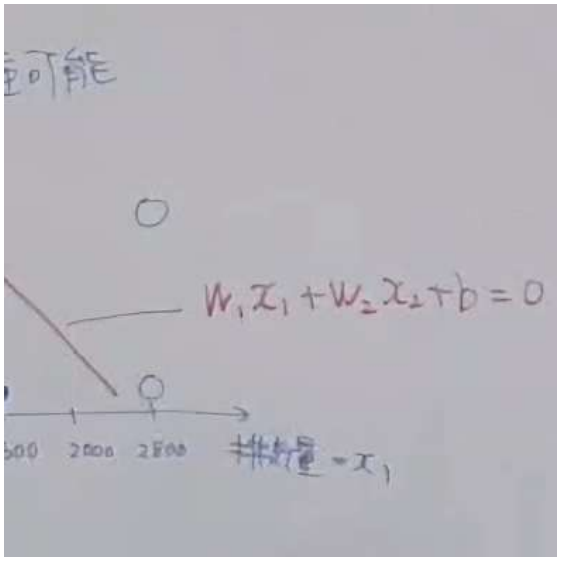
6. 読みやすい（　　，　　，　　，　　）読みにくい



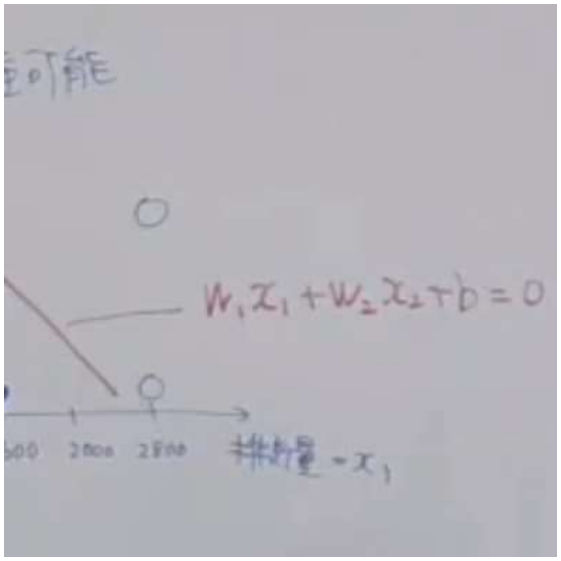
A



B

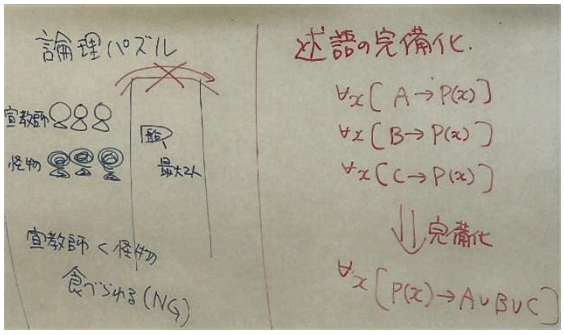


C

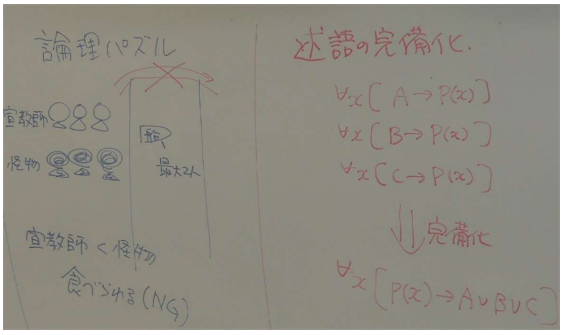


D

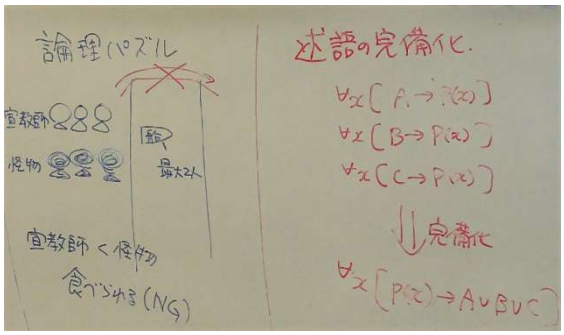
7. 読みやすい（　　,　　,　　,　　）読みにくい



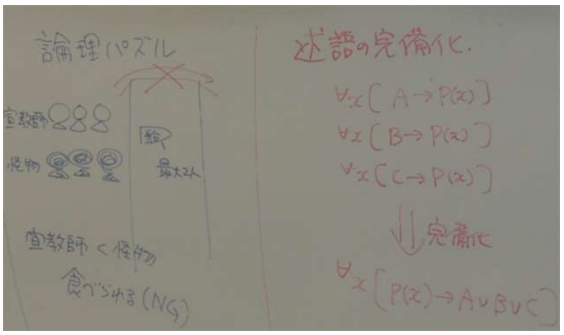
A



B

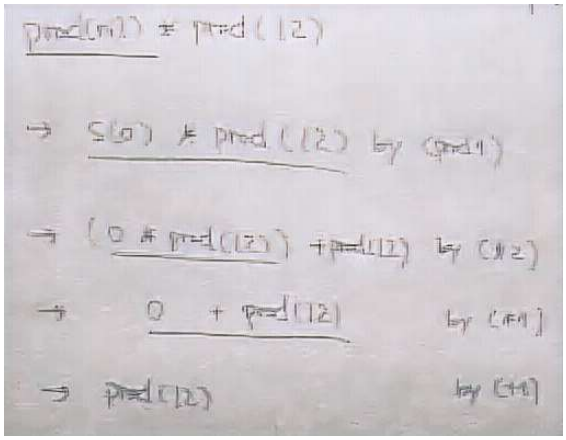


C

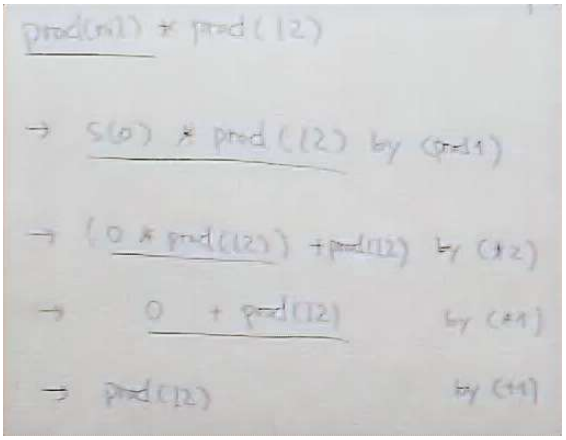


D

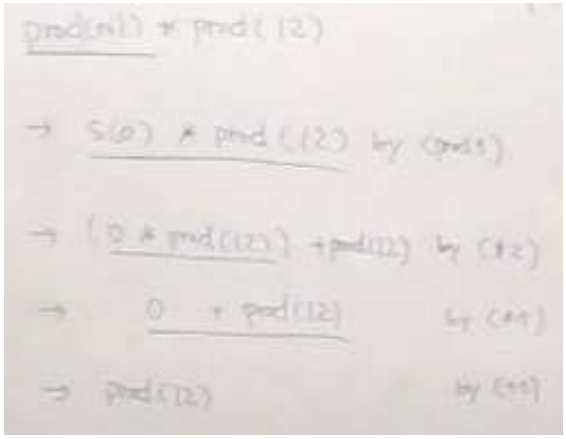
8. 読みやすい（　　,　　,　　,　　）読みにくい



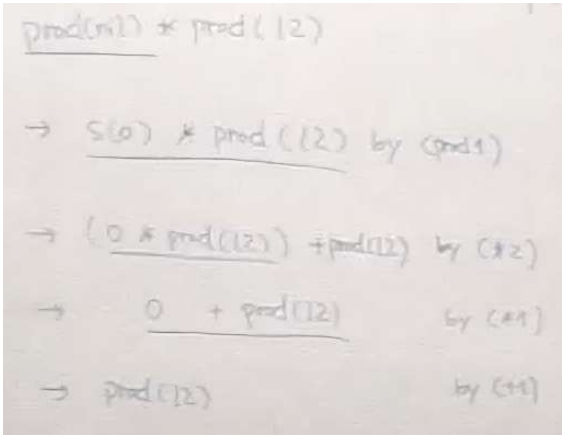
A



B



C



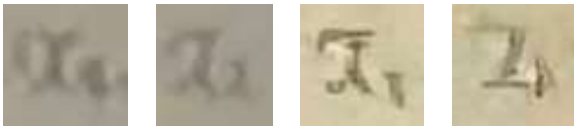
D

客観評価

以下の画像の内容を書いてください。Please write the content of the following pictures.

答え:＿＿＿＿，＿＿＿＿，＿＿＿＿，＿＿＿＿。

答え:＿＿＿＿，＿＿＿＿，＿＿＿＿，＿＿＿＿。



答え:＿＿＿＿，＿＿＿＿，＿＿＿＿，＿＿＿＿。

答え:＿＿＿＿，＿＿＿＿，＿＿＿＿，＿＿＿＿。

答え:＿＿＿＿，＿＿＿＿，＿＿＿＿，＿＿＿＿。