情報処理学会論文誌 Vol.62 No.6 1358–1367 (June 2021)

推薦論文

アモーダル補完を応用した文字型CAPTCHA

上妻 拓也[1] 梅澤 猛[2,a)] 大澤 範高[2]

受付日 2020年5月7日, 採録日 2021年3月2日

概要：視覚の補完機能であるアモーダル補完を応用し，人間には負担が大きすぎず，自動文字認識には攻撃コストが増加し難度が高い CAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）を生成する手法を提案し，評価を行った．提案手法では背景色の図形によって文字の一部を欠けさせた欠損画像と，欠損画像上の欠けた部分を覆い隠すことができる図形を透明背景に描画したマスク画像の 2 枚を提示する．欠損部にマスクがかかるように 2 枚の画像を重ね合わせると，アモーダル補完の効果により人間にとっては容易に文字認識ができる．画像の重ね合わせ操作は，ボットが攻撃に必要とするコストを増大させると期待できるが，欠損画像のみから文字認識されると効果がない．そこで，3 種類の認識困難化手法を組み合わせた欠損画像を生成し，畳み込みニューラルネットワークによる自動文字認識率を評価した．反転ノイズ重畳と文字幅・間隔不均一化の組合せが最も効果があり，正解率を認識困難化手法を使わない場合の 0.946 から 0.788 に低減可能であることを示した．また，被験者による文字列読み取り実験によって，提案した CAPTCHA に対する正解率と解答時間および操作に対する主観的な負荷について調査した．画像の重ね合わせを適切に行うことで，アモーダル補完の効果によって文字の認識が容易になることを確認し，画像の重ね合わせ操作に対する負担感軽減が必要であることも明らかにした．

キーワード：CAPTCHA，文字認識，機械学習，アモーダル補完，CNN

Character-based CAPTCHA Using Amodal Completion

Takuya Kozuma[1] Takeshi Umezawa[2,a)] Noritaka Osawa[2]

Received: May 7, 2020, Accepted: March 2, 2021

Abstract: This paper proposes and evaluates a method to generate character-based CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) using amodal complement which is a visual complement function. They are not too difficult for humans to solve but demanding for automatic character recognizers. The proposed method presents a partially visible image, referred to as a partial image, and an occluder image, referred to as a mask image, for CAPTCHA. The partial image is an image of characters overlapped with objects of background color. The mask image has objects, which can hide invisible parts of the partial image, on a transparent image. By appropriately superposing the mask image on the partial image, humans can easily read characters on the combined image. Superposition can increase the cost of attacks if the partial image cannot be recognized by a recognizer. Therefore, this paper introduces three types of image modification methods that decrease an automatic character recognition rate of partial images, and evaluates the recognition rate by the convolutional neural network. The results show that a combination of inversion noise superposition and non-uniform character width/space is the most effective. Another experiment is conducted to evaluate human performance and aspects of the proposed CAPTCHA. The results of the experiment confirm that appropriately superposed images can be easily recognized, and also clarifies that it is necessary to reduce workload for superposing the two images appropriately.

Keywords: CAPTCHA, character recognition, machine learning, amodal complementation, CNN

[1] 株式会社コナミデジタルエンタテインメント
Konami Digital Entertainment Co., Ltd., Chuo, Tokyo 104–0061, Japan
[2] 千葉大学
Chiba University, Chiba 263–8522, Japan
[a)] ume@chiba-u.jp

本論文の内容は 2019 年 7 月のマルチメディア，分散，協調とモバイル（DICOMO2019）シンポジウムで報告され，セキュリティ心理学とトラスト研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

1. はじめに

自動化されたプログラムであるボットによるフリーメールアドレスの不正取得，ブログやインターネット掲示板へのスパムコメントの投稿・書き込み，オンラインチケットの不正購入が問題となっている．この問題を軽減するために Web サービス利用者が人間かボットかを判別する反転チューリングテストの一種である CAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）[1], [2] が用いられている．その中でも画像中に含まれる文字列を解答する文字型が現在広く利用されているが，機械学習技術を応用した自動文字認識攻撃により，突破することが可能であることが示されている [3], [4], [5], [6]．文字型 CAPTCHA 画像を複雑化させる対策が多くとられているが，深層学習を利用した文字認識に対しては効果が少なく，人間の負担が増すだけになるという問題がある．

本研究では，人間の視覚の機能であり，錯視の一種である，対象の一部が隠されて見えない場合に欠けた部分が補完されるアモーダル補完を応用し，人間には負担が大きすぎず，自動文字認識には攻撃に必要となる計算コスト（攻撃コスト）を増加させ，誤答率として測定できる耐性（攻撃難度）が高い CAPTCHA を提案しその評価を行った．提案手法では，背景色の図形によって文字の一部を欠けさせた欠損画像と，欠損画像上の欠けた部分を覆い隠すことができる図形を透明背景上に描画したマスク画像の 2 枚を提示する．欠損部の上にマスクがかかるように 2 枚の画像を適切に重ね合わせると，アモーダル補完の効果により人間にとっては文字認識が容易になる．従来の文字型 CAPTCHA にはない画像の重ね合わせによってボットが調べなければならない重ね合わせの組合せを増加させ，攻撃コストを増加させる．欠損画像のみからの文字認識の困難化を目的とする 3 種類の手法を提案・評価し，攻撃難度を高めるための有効性を明らかにした．また，提案手法が解答者に要求する画像の重ね合わせ操作の負担や正解率を評価するために，被験者による文字列読み取り実験を行った．

2. 関連研究

CAPTCHA [1] はサービス利用者にテストを出題し，解答の内容によって人間とボットの判別を行う．このテストは人間にとっては容易で，ボットに対しては困難となるように設計されている．CAPTCHA の例として次の 4 種類があげられる [2]．

(1) 文字型

文字列を含む画像を提示し，画像中の文字列を解答させる．

(2) 分類型

グループ分けしたいくつかの画像群と，グループ分けされていない 1 枚の画像を表示する．画像群からグループ分けの法則を見つけ，グループ分けされていない 1 枚の画像がどのグループに属すかを解答させる．

(3) 画像型

大規模な画像データベースから画像をピックアップして提示し，画像中に写っている物体の名称を解答させたり，指定された物体が写っている画像を解答させたりする．

(4) 音声型

単語や何桁かの数字を発話している音声を歪ませたものを再生し，内容を解答させる．

本研究では，このなかで最も広く利用されている文字型 CAPTCHA を対象とした．

2.1 文字型 CAPTCHA

文字型 CAPTCHA は文字列を含む画像を提示し，画像中の文字列を解答させることで人間とボットの判別を行う．文字型 CAPTCHA は画像に描画する文字，文字数，文字の並びなどを変化させることで様々な文字列画像を容易に生成することができるため，生成コストが小さい．また，世界中の多くの人が扱うことのできるアルファベットや数字を利用することで，グローバルな環境にも対応可能であることから，広く利用されている．

歪みやノイズを加えた画像内の文字を自動認識することは困難であったため，ボット対策として利用されてきたが，光学文字認識（OCR）技術の発展や機械学習の応用による自動文字認識の精度向上にともない，既存の多くの文字型 CAPTCHA がボットに対して脆弱であることが示されている [3], [4], [5]．

2.2 文字型 CAPTCHA に対する攻撃

ボットによる文字型 CAPTCHA に対する攻撃は，前処理，領域分割，単一文字認識の 3 つのステップで構成されてきた．前処理ではテキストエリアの切り出し，回転，ノイズの除去，二値化などを行い，領域分割と単一文字認識の精度を高める．領域分割では CAPTCHA 画像をそれぞれが単一の文字を含む小さい領域へ分割する．単一文字認識では分割した単一文字画像から機械学習を利用して文字の認識を行う．

Convolutional Neural Network（CNN）による単一文字認識精度は人間の文字認識精度を上回っているため，文字型 CAPTCHA の攻撃耐性は領域分割の難度に依存している [3], [4]．そのため，Hollow CAPTCHAs [7]，Connecting Characters Together（CCT）[8], Two-Layer CAPTCHA [9] などの様々な領域分割への対策を施した CAPTCHA が使用されているが，それぞれの対策を無効化する前処理や深層学習を組み合わせた手法が提案されており，高い精度で領域分割が可能になり，文字認識を行われてしまうことが明らかとなっている [5], [6]．

さらに，Hu らは，前処理と領域分割の画像処理を分けて行う従来の方法ではなく，CNN を用いて CAPTCHA 画像から文字列を認識する攻撃手法を提案している [10]．前処理と領域分割を分けて行う従来の方法では，攻撃対象の CAPTCHA が使用している領域分割対策によって適用する画像処理を切り替える必要があり，未知の領域分割対策に対して有効な画像処理やアルゴリズムを新たに開発する必要があった．Hu らの手法では前処理と領域分割の画像処理を分けずに，CAPTCHA 画像を直接 CNN への入力とし，前処理と領域分割の過程も CNN に学習させることでこれを回避している．5 文字からなる文字列を含む画像に，ノイズや歪みを加えて生成した CAPTCHA 画像データセットによる評価において，前処理と領域分割を行う従来の手法による文字認識精度 0.95 に対し，0.965 の精度であったと報告されている．本研究では新しい文字型 CAPTCHA の提案・評価を行うが，提案した CAPTCHA は従来の文字型 CAPTCHA と異なる点が多く，既存の前処理や領域分割を適用することが難しいことから，Hu らの手法を評価に用いた．

2.3 アモーダル補完を応用した文字型 CAPTCHA

OCR や機械学習による文字認識攻撃への耐性を高める手法として，人間特有の性質であるアモーダル補完を CAPTCHA に応用する手法が提案されている．森らは，アモーダル補完を動画 CAPTCHA に応用している [11]．この研究では，文字画像，ノイズ画像，円形遮蔽物の組合せがフレームごとに変化する動画 CAPTCHA を作成した．画像は 6×6 のパネルからなり，フレームごとに 1 個から 2 個のパネルが入れ替わる．ユーザはノイズと遮蔽物が適切な位置に重なるフレームのときにアモーダル補完によって文字認識が可能となる．遮蔽物が適切な位置に存在しない場合は文字を知覚することができない．ボットがこの動画 CAPTCHA を突破するためにはフレームごとに画像を解析する必要があり，攻撃コストが大きい．一方で，人間が見え方を制御できず，文字認識が可能となる画像が表示されている時間はごくわずかであり，人間に対する負担も大きいという問題がある．

本研究では，静止画像を使用する従来の文字型 CAPTCHA にアモーダル補完を応用することで，人が見え方を制御できるようにし，文字認識が可能となる時間の制約を取り除き，先行研究のアモーダル補完を利用した動画 CAPTCHA よりも人間に対する負担を小さくすることを目指す．

3. アモーダル補完を応用した CAPTCHA

本研究では，背景色の図形によって文字の一部を欠けさせた欠損画像と，欠損画像上の欠けた部分（欠損部）を覆い隠すことができる図形（遮蔽部）を透明画像に描画したマスク画像の 2 枚を提示する手法を提案する．解答者は欠損部が遮蔽部で覆われるように，欠損画像上の適切な位置にマスク画像を重ね合わせ，合成画像を作成することで文字認識が可能となる．欠損画像，マスク画像の例をそれぞれ**図 1** (a)，(b) に示す．欠損画像は背景色の図形によって文字が欠けているため，画像中の背景色部分が背景か欠損部かの判別がつかない．これにより，どの断片どうしが 1 文字を構成しているかの把握が困難になり，文字認識を行うことが難くなっている．一方，マスク画像にはマスクのみが描画され，文字の情報を含まない．そのためマスク画像から文字認識を行うことも不可能である．

(a) 欠損画像　(b) マスク画像

図 1 欠損画像とマスク画像の例
Fig. 1 Examples of masked image and mask-image.

図 2 合成画像の例
Fig. 2 An example of synthetic image.

合成画像の例を**図 2** に示す．人間はアモーダル補完の効果により，マスクによって遮蔽された部分を補完することで文字認識が容易になる．ボットもマスクによって背景と欠損部が明確になり，合成画像から文字認識を行うことが可能となるが，ボットは合成画像を作成するために画像の重ね合わせを総あたり的に行い，作成したすべての合成画像に対して文字認識を行う必要があると考えられるため，欠損画像だけで攻撃が成功しにくければ，攻撃コストを増加させることができる．

たとえば，欠損画像とマスク画像の位置合わせを行う際に，マスク画像を縦 H pixel，横 W pixel の移動可能範囲でずらすことが可能で，R 種類の回転が可能な場合には，$H \times W \times R$ 通りの重ね合わせの組合せがあり，すべての組合せの合成画像に対して文字認識をする攻撃のコストは移動可能範囲と回転可能種類を増やすことで増加させることができる．

提案手法は，ボットよりも人間が速く，高い精度で解答できるようにすることを目的とするものではなく，人間が解答できる問題に対するボットの攻撃コストを増加させることを目的としている．また，画像操作による攻撃コスト増加が意味を持つために，画像操作を必要としない欠損画像に対する攻撃難度を高める．

4. 認識困難化手法

欠損画像のみから文字認識が可能であれば，ボットが画像の重ね合わせを行う必要がなく，攻撃コストを増加させ

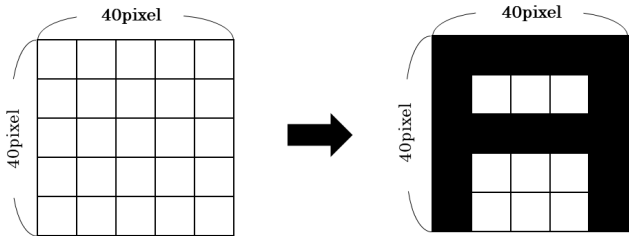


図 3 文字画像の作成例（例：‘A’）
Fig. 3 An example of generating a character image (e.g.: ‘A’).

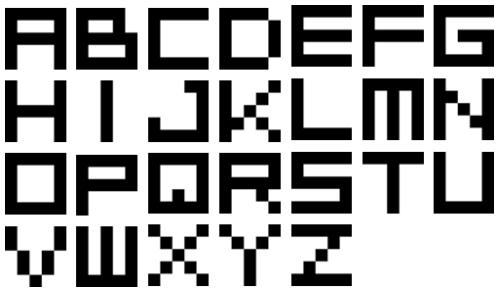


図 4 作成した文字画像
Fig. 4 A set of generated characters.

ることができない．そこで，欠損画像のボットに対する攻撃耐性を検討する必要がある．

筆者らのこれまでの実験において [12]，欠損画像に多くの文字色のノイズを重畳しても機械学習による認識の正解率は低下しなかった．これは，欠損画像に文字認識に必要な情報がある場合には，それに単純な文字色ノイズを重畳しても認識正解率を低下させることができないことを意味している．また，前述のように Hu らの研究 [10] においても，文字にノイズや歪みを加えても正解率 0.965 で 5 文字を認識できている．これらのことから，欠損画像上に適切なマスク画像を重ねて生成される，欠損がなくノイズの重畳された画像は機械学習によって高精度で認識が可能である．

単純な文字色ノイズの重畳には効果がないことから，欠損画像に反転ノイズ重畳，無効文字挿入，文字幅・間隔不均一化という 3 種類の手法の組合せを加えることで認識を困難にできるかを検討する．以下ではその評価のための準備として，評価実験の条件と各認識困難化手法を説明する．

4.1 文字列画像

評価実験には 1 マスが 8 × 8 pixel で構成される 5 × 5 マスの縦横 40 × 40 pixel の領域を占める二値の文字画像を用いた．領域を縦横 5 等分した 5 × 5 マスを使用して（**図 3**），アルファベットの大文字 26 文字を用意した（**図 4**）．

4.2 欠損画像

文字列画像に欠損個所をつくる欠損部形状は，対象となる文字画像内（**図 5** (a)）で無作為に選択した座標（図 5 (b)）を中心とする 16 × 16 pixel の正方形（図 5 (c)）とした．確実に欠損個所が生じるよう，文字色の座標のみを選択対象とした．また，欠損個所が偏ることを避けるため，背景色図形は互いに 6 pixel 以上の間隔を空けることとした．

(a) 選択可能範囲 (b) 座標の選択 (c) 欠損の生成

図 5 欠損画像の作成例
Fig. 5 Examples of generating masked image.

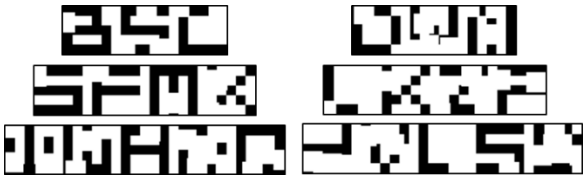
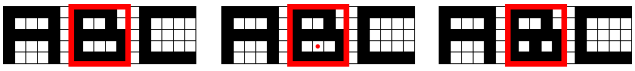


図 6 欠損画像の例
Fig. 6 Examples of masked image.



(a) 選択可能範囲 (b) 座標の選択 (c) ノイズの重畳

図 7 反転ノイズの重畳例
Fig. 7 An example of generating image with inverse noise.

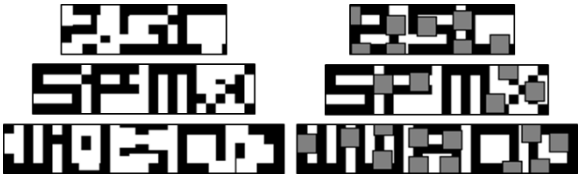


図 8 反転ノイズを含む欠損画像と対応する合成画像の例
Fig. 8 Examples of masked image with inverse noise.

文字色部分の少ない文字（例：‘I’）と多い文字（例：‘W’）の欠損が同程度となるように，使用する欠損部形状の数は 1～（文字の黒画素数/256）の範囲から無作為に決定した．欠損画像の例を**図 6** に示す．

4.3 反転ノイズ重畳

対象となる文字画像内で無作為に選択した座標を含むマスについて，文字色と背景色を反転させることでノイズの重畳を行った．**図 7** (a) に反転ノイズ重畳の対象文字が ‘B’ だった場合の座標選択範囲を赤枠で表している．図 7 (b) の座標が選択された場合，座標の画素は白なので，図 7 (c) に示すように選択された座標を含むマスを黒で塗る．ノイズを含む欠損画像と，対応する合成画像の例を**図 8** に示す．

4.4 無効文字挿入

CNN を用いた単一文字認識は，欠損やノイズを含む文字であっても非常に高い精度を示す．そこで，背景色の文

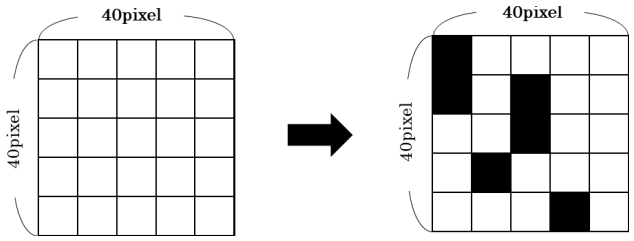


図 9 無効文字の作成例
Fig. 9 An example of generating invalid character image.

図 10 無効文字挿入の例
Fig. 10 An example of inserting invalid character.

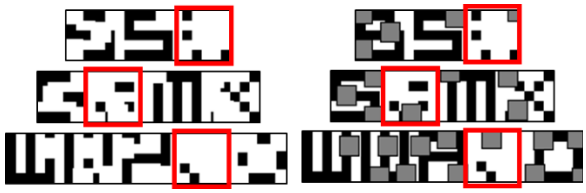


図 11 無効文字と反転ノイズを含む欠損画像と対応する合成画像の例
Fig. 11 Examples of masked image with invalid character and inverse noise.

字画像領域内の複数のマスを文字色にすることで生成した無効な文字の挿入が認識精度に与える影響を調べた．文字画像の 5×5 マスから 3〜7 マスをランダムに選択して文字色とすることで無効文字を作成した（**図 9**）．ただし，人間が文字の一部だと誤認識してしまうのを避けるため，同じ列・行で文字色とするのは 3 マス未満という制約を設けた．文字列中に含まれる 0〜1 文字を無作為に選択し無効文字と入れ替えた．**図 10** は無効文字を文字列に追加する例を表し，「ABC」という文字列画像の ‘B’ を無効文字と入れ替えている．以上の条件で生成した反転ノイズを含む欠損画像と，対応する合成画像の例を**図 11** に示す．

4.5 文字幅・間隔不均一化

文字サイズが均等であると学習モデルによる文字分割が容易であると考えられるため，文字幅や文字の間隔を不均一とすることで認識精度を低下させられる可能性について調べた．

文字画像の作成時のマスを単位とし，列ごとに {2, 4, 6, 8, 10, 12} pixel の中から無作為に幅を変更した（**図 12**）．ただし，文字のバランスが大きく崩れるのを避けるため，対象とする列は文字の横線が描かれた列と文字間のみとした．図 12 の赤枠で囲まれた領域が変更対象の列を示す．A，B，C の文字のぞれぞれの横線が描かれた 1 領域と文字間の空白領域が対象として選択されている．文字・文字

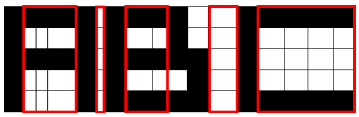


図 12 文字幅・間隔不均一化の例
Fig. 12 An example of ununiformed character width and spacing.

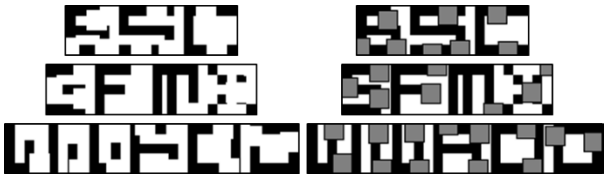


図 13 文字幅・間隔を不均一化した欠損画像と対応する合成画像の例
Fig. 13 Examples of masked image with ununiformed character width and spacing.

間サイズを不均一化した欠損画像と，対応する合成画像の例を**図 13** に示す．

5. CNN による認識実験

4 章で説明した認識困難化手法の組合せによって，CNN による欠損画像に対する攻撃難度が高まることを実験的に評価した．Hu らの自動文字列認識攻撃手法に基づき CNN による認識実験を行った．アルファベット大文字 26 文字の中から重複を許す 3〜5 文字を無作為に選択して文字列を作成し，認識困難化の組合せごとに学習用データ 10,000 件，検証用データ 2,000 件，テストデータ 1,000 件を合わせてデータセットを構成した．

認識モデルはつねに 5 文字を出力する．認識文字列が 3 文字もしくは 4 文字の場合には，後にパディング文字を出力する．また，各文字のクラス数はアルファベット大文字 26 種類，無効文字，パディング文字の計 28 である．

5.1 CNN 構造

CNN の実装には，深層学習フレームワークである TensorFlow [13] をバックエンドとしたライブラリである Keras [14] を用いた．CNN のモデルは，Hu らによって提案された文字型 CAPTCHA 認識用モデルと同じである．これは中間層に畳み込み層 10 層，プーリング層 4 層，全結合層 3 層を含む全 17 層からなるディープニューラルネットワークである．4 番目のプーリング層の後，モデルは 5 つに分岐し，各出力がそれぞれ文字列中の 1 文字に対応している．

白黒の二値画像を縮小すると歪みを生じ，文字など形状が変形する．画像の変形を図るとともにモデルの学習時間短縮を図るために，画像の縦横を半分にリサイズし，コントラスト正規化を行ったものを CNN への入力とした．バッチサイズは 128，エポック数は 300 とした．学習係数

表 1 認識困難化の組合せに対する CNN の正解率
Table 1 The accuracy rate for each condition.

無効文字挿入	文字幅・間隔不均一化	反転ノイズ重畳	正解率
—	—	—	0.946
—	—	✔	0.914
—	✔	—	0.917
—	✔	✔	0.788
✔	—	—	0.959
✔	—	✔	0.913
✔	✔	—	0.912
✔	✔	✔	0.830

の最適化には SGD（Stochastic Gradient Descent）を使用し，学習係数は 0.001 とした．なお，ドロップアウトはすべての全結合層後に行い，ドロップアウト率は 0.5 とした．

5.2 実験結果

認識困難化手法として反転ノイズ重畳，無効文字挿入，文字幅・間隔不均一化を組み合わせた条件で生成したデータセットの学習データで訓練を行った CNN モデルによるテストデータに対する正解率を**表 1** に示す．ここで，欠損文字列画像に含まれる，無効文字を含むすべての文字を正しく認識した場合を正解とした．たとえば，認識対象の文字列が 4 文字の「A」「無効文字」「B」「F」という列の場合は，「A」「無効文字」「B」「F」「パディング文字」という列の出力が得られた場合のみ正解である．

1 文字でも誤認識をした場合には不正解とした．正解率は，テストデータ N 件のうち，正解の件数 R の割合（R/N）である．

反転ノイズ重畳，文字幅・間隔不均一化は個別に行うことでも正解率はそれぞれ 0.914，0.917 であり，まったく認識困難化を行わない場合の正解率 0.946 と比べて低減できている．また，反転ノイズ重畳と文字幅・間隔不均一化を組み合わせることで正解率は 0.788 と最も低くなり，それぞれの手法を個別で行うよりも正解率を低下させることができることから，有効な手法であることが示唆される．

3 つの手法すべてを適用した場合には，正解率が 0.830 であり，ノイズ重畳と文字幅・間隔不均一化の組合せに無効文字挿入を加えても正解率を低下させる効果が見られなかった．また，無効文字挿入だけを行った場合には正解率が 0.959 と最も高くなった．

5.3 考察

Hu らは，文字型 CAPTCHA に対して CNN を用いて 0.965 の正解率を示している．それに対し，本実験の条件では前記のとおり，同じ CNN モデルを使用して 0.788 の正解率を示した．このことから，提案手法における欠損画像の文字認識攻撃に対する耐性は従来の文字型 CAPTCHA よりも高く，提案手法が従来手法よりも文字認識攻撃に対してより高い耐性，すなわち，高い攻撃難度を有している．

ただし，ボットの突破率が 1%以下であることを要求する考え方もあり [15]，欠損画像に対する自動認識精度をさらに低下させる認識困難化手法や他のボット耐性を持つ手法と組み合わせた統合手法が今後の課題となる．

文字列全体の正解率に加えて，文字ごとの認識率を確認した．認識困難化なしの場合には，誤認識は ‘A’ と ‘P’ の 2 つの文字で多く発生しており，他の文字は高い精度で文字認識が行われていた．‘A’ を ‘P’ に，‘P’ を ‘F’ に多く誤認識していた．これは ‘A’ と ‘P’ の構造が似ており，‘A’ の右下部分を欠損させることで ‘P’ と同じ画像になるためと考える．‘P’ と ‘F’ も同様である．一方，CNN はパディング文字を認識率 1.00 で認識していることから，文字数を正確に推定できていると考えられる．文字型 CAPTCHA の攻撃耐性を向上させる手法として，文字数を可変にすることは一般的であるが，本実験では文字サイズが固定であり，CNN への入力の際にデータセット内の最大画像サイズに合わせて画像をパディングしていることから，文字数を推定することは容易であった．そのため，文字数を可変にすることは，分類が簡単なパディング文字というクラスを追加するだけになっていると考える．

反転ノイズ重畳を行った場合には，‘A’ と ‘P’ 以外の文字に対する認識率が認識困難化をまったく行わない場合の認識率以下であった．特に，‘F’ を ‘P’，‘O’ を ‘Q’，に誤認識するケースが増加した．これは，これらの文字の構造に 8×8 pixel の 1 マス分の違いしかなく，ノイズが描画される位置によって，‘F’ は ‘P’，‘O’ は ‘Q’ とまったく同じ画像になるためである．誤認識は，構造が非常に似ている文字どうしの間で発生していることが多く，特徴的な構造である ‘K’，‘N’，‘X’，‘Y’，‘Z’ などの文字は誤認識が起こっていない．このことから構造が特徴的な文字を誤認識させるためには，誤認識を誘導できるよう反転ノイズを重畳する位置を決定することが考えられる．

文字幅・間隔不均一化を施した場合の文字ごとの認識率を認識困難化なしの場合と比較すると，誤認識が発生する文字の種類が増えていた．また，反転ノイズ重畳の結果と異なり，多少の偏りはあるものの，様々な文字の間で誤認識が発生していることが確認できた．

無効文字挿入を行った際の無効文字を含む欠損画像に対する文字ごとの認識率は 0.996 であり，他の文字に誤認識したのは 2 件，他の文字を無効文字に誤認識したのは 1 件だけであった．したがって，無効文字を追加することは学習モデルにとって，分類が容易な 1 つの文字が増えることになり，全体の認識率が高くなったと考えられる．ただし，本実験では無効文字の生成方法に，3〜7 個のマスを黒で塗る，マス目の同じ列・行には 3 マス以上描かないという制約を設けている．この制約によって，無効文字が他の文字

とは異なる大きな特徴を持っている可能性がある．効果的な制約について検討することが今後必要である．

6. 被験者による認識実験

ボットの影響を低下させるために攻撃コストを増やせても，人間が解答できない問題ではCAPTCHAとしての意味がない．提案手法は，画像の重ね合わせを被験者に要求するため，操作負荷や誤解答の増加が懸念される．そこで，それらの増加の程度を評価し，人間が現実的な時間内に解答可能な問題を生成するために考慮すべき点を明らかにするために，被験者によるCAPTCHA読み取り実験を行った．

6.1 実験条件

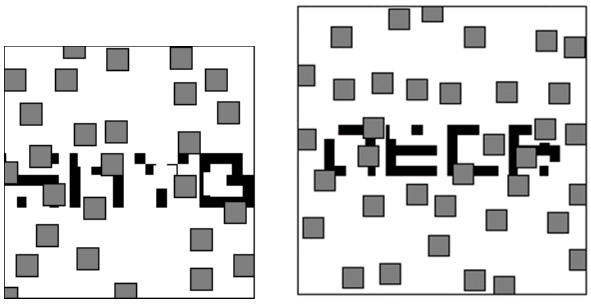
欠損画像とマスク画像を提示し，画像の重ね合わせを行うことができる実験用アプリケーションをPC上に作成した．被験者は，マウスまたはキーボードの方向キーによって画像の重ね合わせ操作を行うことができる．操作方法の違いによる影響を調べるため，操作条件を変えたタスクを用意した．被験者は大学学部生および大学院生合わせて6名であった．被験者にはまず今回使用する文字画像の一覧を提示し，各アルファベットがどの文字画像に対応するかを確認してもらった．

タスクの種類と欠損画像の生成手法の組合せから実験条件が決まる．欠損画像の生成手法の組合せから，問題セットを生成した．また，タスクにおける問題セットは各被験者で異なる順序で提示した．タスクが終了するごとに被験者に対し「本タスクで提示されたCAPTCHAの解答負荷は重い」にLikert尺度で，1（強く不同意）～5（強く同意）の5段階で回答する質問紙調査を行った．

6.2 タスク

タスクは次の4種類を課した．1）欠損画像のみを提示し，マスク画像を使わずに読み取りを行う，2）欠損画像に対して，あらかじめ適切な位置にマスク画像を重ねた状態で読み取りを行う，3）欠損画像の上下移動とマスク画像の回転で画像を重ね合わせて読み取りを行う，4）欠損画像の上下左右移動で画像を重ね合わせて読み取りを行う．

タスク3とタスク4の違いは，回転操作の有無と平行移動の範囲である．**図 14** にタスク3とタスク4の画像例を示す．タスク3においては，欠損画像の横幅を一辺の長さとする正方形のマスク画像を用い，マウスによるドラッグおよびホイールスクロールと，キーボードの上下キーを使って欠損画像の上下移動を，マウスの右クリックと，キーボードの左右キーを使ってマスク画像を90度回転する．タスク4においては，欠損画像の横幅+40 pixelを一辺の長さとする正方形のマスク画像を用い，マウスによるドラッグおよびホイールスクロールと，キーボードの上下左右キーを使って欠損画像の上下左右移動を行う．



（a）タスク3　　（b）タスク4

図 14 欠損画像とマスク画像の例
Fig. 14 Examples of masked image with mask images.

タスク1～3は各20問，タスク4は10問で構成され，各タスクにおいて半数は全被験者共通の問題，残り半数は被験者ごとに異なる問題とした．

6.3 欠損画像種別

欠損画像の生成手法としては，A）欠損のみ，B）反転ノイズを重畳，C）文字幅・間隔の不均一化，D）反転ノイズを加え，文字幅・間隔を不均一化する4種類を選択した．無効文字挿入は，CNNによる認識実験から自動文字認識攻撃耐性への貢献が明確でないことから条件に含めなかった．

6.4 実験結果

図 15 に，各問題セットにおける各タスクについての平均解答時間を示す．エラーバーで，最大値と最小値を示す．移動回転操作が不要であることから予想されるとおり，タスク2に対する平均解答時間が最も短くなり，被験者間でも差があまり生じていないことが分かる．対して，画像の重ね合わせを行うタスク3とタスク4ではタスク2の倍以上の平均解答時間となり，被験者間で大きく差が生じていることが分かる．

図 16 に，各実験条件における平均正解率を示す．エラーバーで最大値と最小値を示す．平均解答時間と同様に，タスク2に対する平均正解率が最も高くなり，被験者間で差があまり生じていない．対して，画像の重ね合わせを行うタスク3とタスク4では平均正解率が低くなるものの，タスク1よりは高い値を示した．

図 17 に，各実験条件に対する質問紙調査結果を示す．画像の重ね合わせを行うタスク3とタスク4に対しては解答負荷が重いと答える被験者が多かったことが分かる．

6.5 考察

被験者による認識実験の全問題の平均解答時間と正解率について考察する．欠損画像のみを表示する条件1-Aの正解率が0.425であるのに対し，合成画像を表示する条件

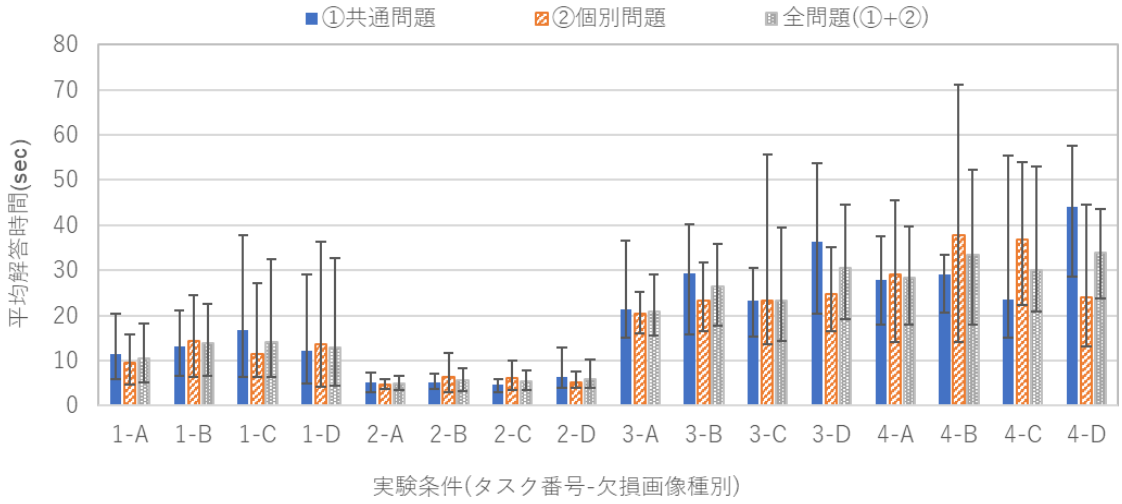


図 15 文字列読み取り時間

Fig. 15 Average response time for each condition.

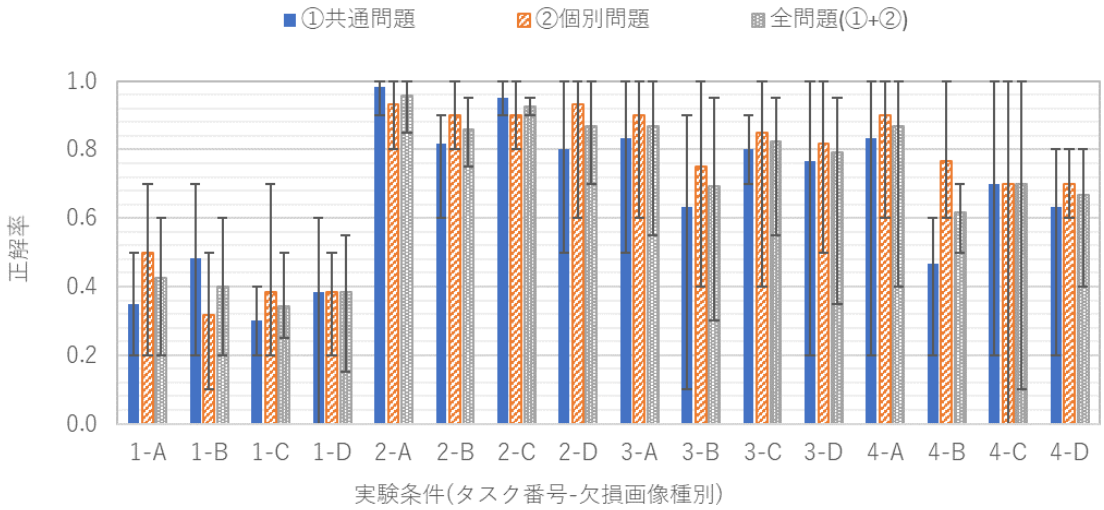


図 16 文字列読み取り正解率

Fig. 16 Correct answer rates for each condition.

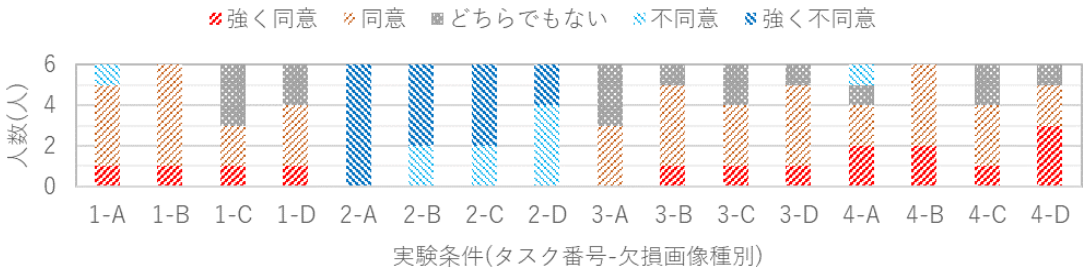


図 17 操作負荷についての回答（Q：本タスクで提示された CAPTCHA の解答負荷は重い）

Fig. 17 Result of questionnaire on workload.

2-A の正解率は 0.958 と高い．また，条件 2-A に対する平均解答時間は 4.947 秒であり，条件 1-A の 10.567 秒の半分以下であることが分かる．さらに，質問紙調査結果からも解答負荷が重いと感じた被験者，すなわち，強く同意もしくは同意と回答した被験者の割合は条件 1-A では 83.3%であったのに対して，条件 2-A に対して解答負荷が重いと感じた被験者の割合は 0%であった．

全問題を共通問題と個別問題に分けた場合の平均解答時間や正解率に着目しても，欠損画像生成種別の B，C，D においても同様の傾向が見られる．以上のことから人間は欠損部分をマスクで隠すことによって起こるアモーダル補完の効果によって，合成画像からであれば，文字を認識しやすいということが確認された．

一方，どの欠損画像種別においてもタスク 3，タスク 4

はタスク 2 よりも正解率が低くなっており，平均解答時間の点でも，タスク 3，タスク 4 は条件 3-A の被験者個別問題に 20.443 秒以上の解答時間を要し，被験者によっては 1 問に 1 分以上かかる例も見られた．タスク 3，タスク 4 において，画像の重ね合わせを適切に行うことができれば，タスク 2 の合成画像と同じ画像を作成することができるため，タスク 2 とタスク 3，タスク 4 の間に正解率と解答時間の差が生じたのは，画像の重ね合わせを適切に行うことが難しく，適切な合成画像を作成することができなかったからであると考える．質問紙調査でも，タスク 3 とタスク 4 に対して解答負荷が重いと感じた被験者は 50%を超えており，本実験で行った画像の重ね合わせ方法は被験者が負荷を感じやすいことが示唆された．

被験者による認識実験の結果から，タスク 3，タスク 4 のような画像の上下移動・回転や 4 方向移動を必要とする操作方法は，被験者に負荷を感じさせやすく，解答時間が長くなり，正解率にも影響が出ることが示された．提案手法は先行研究の動画 CAPTCHA [11] と異なり画像の重なりを操作できるが，画像の重ね合わせの方法をより簡単化するなど，被験者の負担を低減する方法についてさらに検討を行う必要がある．ただし，タスク 3 では縦に 120 pixel 移動可能であり，回転による向きが 4 種類あるため，ボットの総あたり攻撃コストを 1 画像に対する攻撃コストの 480 倍にでき，タスク 4 では縦に 160 pixel，横に 40 pixel の移動範囲があるため，総あたり攻撃コストを 6,400 倍にすることができるが，画像の重ね合わせパターン数を減少させると，ボットの攻撃コストが低下する問題がある．

ボットの攻撃コスト低下の問題を緩和するために reCAPTCHA [16] v2 や v3 などでの操作履歴による判別のように，アモーダル補完を得るためのマスク画像操作の履歴を人間とボットの区別に利用することができる．アモーダル補完と操作履歴を組み合わせて利用することによって，ボットは文字を推定するだけではなく，適切な操作を適切なタイミングで再現する必要があり，攻撃コストを増やすことができる．アモーダル補完を利用した提案手法と操作履歴の併用によって人間の負担を増やすことなく，ボットの攻撃コストを増加させることが期待できる．

7. まとめ

本研究では，人間の持つ視覚の補完機能であるアモーダル補完を応用し，人間には負担が大きすぎず，自動文字認識には攻撃コストが増加し攻撃難度が高い CAPTCHA を生成する手法として，背景色の図形によって文字の一部を欠けさせた欠損画像と，マスク画像の 2 つを提示し，欠損部もしくは隠されるべき部分にマスクがかかるように 2 つの画像を重ね合わせる手法を提案し，評価した．

合成画像を作成可能な重ね合わせの組合せを増やすことで攻撃コストを増加させることができることを示した．また，3 種類の欠損画像生成手法の組合せによって生成した欠損画像に対する CNN 学習モデルによる文字認識精度を評価し，反転ノイズ重畳，文字幅・間隔不均一化によって，正解率を 0.946 から 0.788 に低減できたことを示した．

また，提案手法による画像を用いた被験者による文字列読み取り実験を行い，画像の重ね合わせを適切に行うことができれば，アモーダル補完の効果によって文字の認識が容易になることを確認した．ただし，実験で行った画像の重ね合わせ操作に対して負担感を感じる被験者が多く，負担感軽減や解答時間の短縮を可能とする改良方法の検討が課題であることも明らかにした．

参考文献

[1] von Ahn, L., Blum, M., Hopper, N. and Langford, J.: CAPTCHA: Using hard AI problems for security, *International Conference on the Theory and Applications of Cryptographic Techniques*, pp.294–311, Springer (2003).

[2] von Ahn, L., Blum, M. and Langford, J.: Telling humans and computers apart (automatically) or how lazy cryptographers do AI, *Advanced in Cryptology*, Lecture Notes in Computer Science, pp.294–311 (2003).

[3] Chellapilla, K., Larson, K., Simard, P. and Czerwinski, M.: Computers beat humans at single character recognition in reading based human interaction proofs, *The 2nd Conference on Email and Anti-Spam* (2005).

[4] Yan, J. and El Ahmad, A.S.: A Low-cost Attack on a Microsoft CAPTCHA, *Proc. 15th ACM Conference on Computer and Communications Security*, pp.543–554 (2008).

[5] Hussain, R., Gao, H. and Shaikh, R.A.: Segmentation of connected characters in text-based CAPTCHAs for intelligent character recognition, *Multimedia Tools and Applications*, Vol.76, No.24, pp.25547–25561 (2017).

[6] Tan, M., Gao, H., Zhang, Y., Liu, Y., Zhang, P. and Wang, P.: Research on Deep Learning Techniques in Breaking Text-based CAPTCHAs and Designing Image-based CAPTCHA, *IEEE Trans. Information Forensics and Security*, Vol.13, No.10, pp.2522–2537 (2018).

[7] Gao, H., Wang, W., Qi, J., Wang, X., Liu, X. and Yan, J.: The robustness of hollow CAPTCHAs, *Proc. ACM SIGSAC Conference on Computer Communication*, pp.1075–1086 (2013).

[8] Gao, H., Wang, W., Fan, Y., Qi, J. and Liu, X.: The robustness of 'connecting characters together' CAPTCHAs, *Journal of Information Science and Engineering*, Vol.30, pp.347–369 (2014).

[9] Gao, H., Tang, M., Liu, Y., Zhang, P. and Liu, X.: Research on the security of Microsoft's two-layer Captcha, *IEEE Trans. Information Forensics and Security*, Vol.12, No.7, pp.1671–1685 (2017).

[10] Hu, Y., Chen, L. and Cheng, J.: A CAPTCHA recognition technology based on deep learning, *13th IEEE Conference on Industrial Electronics and Applications* (*ICIEA*) (2018).

[11] 森　拓真，宇田隆哉，菊池眞之：アモーダル補完を利用した動画 CAPTCHA の提案，マルチメディア，分散，協調とモバイルシンポジウム 2011 論文集，pp.1518–1525 (2011).

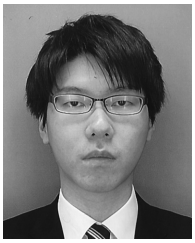
[12] 上妻拓也，梅澤　猛，大澤範高：アモーダル補完を応用した CAPTCHA における文字認識攻撃への耐性評価に関す

る検討，第 16 回情報科学技術フォーラム，L-010 (2017).
[13] Google: TensorFlow, available from ⟨https://www.tensorflow.org⟩ (accessed 2020-09-12).
[14] Keras Google group: Keras Documentation, available from ⟨https://keras.io/⟩ (accessed 2020-05-01).
[15] Bursztein, E., Martin, M. and Mitchell, J.C.: Text-based APTCHA strengths and weaknesses, *Proc. 18th ACM Conference on Computer and Communications Security*, pp.125–138 (2011).
[16] Google: reCAPTCHA, available from ⟨https://www.google.com/recaptcha⟩ (accessed 2020-09-12).

推薦文

DICOMO2019 の発表論文の中で特に評価が高かったため．画像の重ね合わせに着目したアモーダル補完という新しい手法について丁寧な検証を行っており，今後の発展が期待できる．

（セキュリティ心理学とトラスト研究会主査　寺田 真敏）



上妻 拓也

2017 年千葉大学工学部情報画像学科卒業．2019 年同大学大学院博士前期課程修了．在学中は，人間の視覚補完現象を応用した文字型 CAPTCHA 改良の研究に従事．現在，株式会社コナミデジタルエンタテインメントに勤務．

梅澤 猛　（正会員）

2007 年慶應義塾大学大学院開放環境科学専攻後期博士課程修了．博士（工学）．2006 年独立行政法人情報通信研究機構専攻研究員．2011 年千葉大学大学院助教，現在に至る．位置情報システム，ユーザインタフェースに関する研究に従事．IEEE 会員．

大澤 範高　（正会員）

1983 年東京大学理学部情報科学科卒業．1988 年同大学大学院理学系研究科博士課程修了．ソフトウェア開発会社，電気通信大学助手，メディア教育開発センター助教授・教授を経て 2009 年千葉大学大学院教授．理学博士．3 次元ユーザインタフェース，屋内位置推定，システムソフトウェアに興味を持つ．電子情報通信学会，ACM，IEEE 各会員．