PassStyles: 顔特徴の混合によるグラフィカル認証システムの提案

松崎 光[1,a)]　前川 知行[2]　今井 倫太[1]　石井 健太郎[3]

概要：本稿では，StyleGAN によって生成された顔画像を利用したグラフィック認証手法 PassStyles を提案する．グラフィック認証手法はテキスト認証手法と比較してユーザビリティが高いという利点がある．しかし，非正規ユーザが正規ユーザの認証動作を覗き見する，ショルダサーフィン攻撃に弱いという欠点がある．PassStyles は，生成された 9 枚の顔画像から 1 枚の正解画像を選択することによって認証を行う．PassStyles の利点は，ユーザが自分の考えた条件を満たす顔画像をいくつか選択することで，直感的にパスワードを設定できることある．また，StyleGAN によって生成された多様な顔画像と 3×3 の配置により，非正規ユーザが条件を特定することが困難なため，ショルダサーフィンに対する頑健性が高いという利点もある．さらに，PassStyles は StyleGAN のどのレイヤーに特徴が含まれているかを指定する仕組みを持っているため，数枚の画像から多種多様な認証用顔画像を生成することができる．なお，PassStyles は使用する画像を調整するモジュールもあり，生成された画像がユーザの指定する条件に十分に従わないことで発生する認証誤りを減少させる．評価実験の結果，様々な画像の特徴がパスワードの条件となり得ることが示された．さらに，PassStyles はショルダサーフィンに対しても十分安全な認証を行うことができた．

1. はじめに

一般的なパスワード認証においては，非正規ユーザが何かしらの方法で不正にパスワードを取得した場合，正規ユーザ以外が認証を突破出来てしまう．非正規ユーザが認証を突破する方法として，正規ユーザが認証をしている場面を横から覗き見るショルダサーフィンという方法が考えられる．本稿での目的は，ショルダサーフィンに強い認証システムの作成である．

目的を実現するための解決すべき問題として，どのような種類のパスワードを用いるかを考えることが重要である．また，非正規ユーザにとってショルダサーフィンしにくい認証システムであると同時に，正規ユーザにとっても使いやすい認証システムである必要がある．

認証方式の一種として，グラフィカルパスワードがある．グラフィカルパスワードは英数字を用いるテキストパスワードの代わりとして多くの研究がなされている．テキストパスワードにおいては総当たり攻撃や辞書攻撃に弱いという欠点があるが [12] [2]，グラフィカルパスワードにおいては改善されている．また，人間は言語表現よりも画像表現の方が記憶しやすいという特性があり [8]，ユーザビリティにおいてもグラフィカルパスワードは優れている．グラフィカルパスワードは，画像そのものを選択する Cognometric 方式・画像の中の特定の場所を選択する Locimetric 方式・図形の描画操作を行い登録された図形と比較する Drawmetric 方式の 3 つに大きく分類される [3]．本稿で提案する PassStyles は Cognometric 方式であり，既存研究として，ユーザの好みの画像を認証に用いる Awase-E [11]，ランダムアート画像を選択する Déjà Vu [9] や人間の顔画像を選択する PassFaces [10]，および生成図形を選択する石井の研究 [4] が提案されている．

しかしながら，グラフィカルパスワードにおいては，ショルダサーフィンに弱いという欠点がある．特に，スマートフォンなどのタッチ型インタフェースにおいては，ユーザの動作を覗き見ることによって容易にパスワードが漏洩してしまうという欠点がある [1]．ショルダサーフィンに強い認証システムとして，生成図形を用いたワンタイムパスワードを石井の研究 [4] では提案している．しかし，ショルダサーフィンを行った際には，非正規ユーザであっても 27.8%の確率で認証を突破できてしまう．また，石井の研究 [4] では，パスワードを 4 カテゴリ 12 種類しか生成することが出来ない．そのため，パスワードの多様性に欠け，認証システムに慣れるほどショルダサーフィンに対する脆

[1] 慶應義塾大学
[2] 静岡大学
[3] 専修大学
[a)] matsuzaki@ailab.ics.keio.ac.jp

© 2025 Information Processing Society of Japan.

弱性が増すと考えられる．さらに，ユーザがダミー画像を選択したときに，認証に使用する画像を調整する機能も提案されていない．

本稿では, StyleGAN [5] を用いた認証システム PassStyles を提案する．PassStyles においては，StyleGAN によって生成された人間の顔画像を認証に用い，9 枚の顔画像の中から 1 枚の正解画像を選択することを 4 連続で成功させることで認証成功となる．

なお, StyleGAN を単に用いるだけでは，頑健な認証システムを構築できない．認証に用いる顔画像の生成の多様性，非正規ユーザのショルダサーフィンを防ぐ画像配置，誤った画像生成の 3 つの課題に対処する必要がある．1 つ目の課題は，PassStyles を開発するための基礎となるものである．PassStyles は，条件を満たす顔画像と満たさない顔画像の 2 つを混合させることで正解画像とダミー画像を生成するが，正解画像は条件を満たし，ダミー画像は条件を満たさない必要がある．しかしながら，単に StyleMixing を行うだけでは，生成される画像のバリエーションが少なくなる．2 つ目の課題として，非正規ユーザのショルダサーフィンを防ぐことがある．PassStyles を安全な認証システムにするためにショルダサーフィン攻撃への対処は必要不可欠である．仮に，正解の顔画像が他のダミー画像と比較して明らかに異なっている場合，非正規ユーザは容易に正解画像を認識することができる．PassStyles は非正規ユーザが正解画像を選択する事態を避ける必要がある．3 つ目の課題として，誤った画像生成に対処する必要がある．正解画像はすべて条件を満たす画像であり，ダミー画像はすべて条件を満たさない画像であることが理想である．しかし，StyleGAN の性質上，画像生成を完全に制御することはできない．ダミーの顔画像に条件を満たす顔画像が含まれていると，正規ユーザが正解画像を選択できないことがある．従って，PassStyles は画像生成の誤りに対応できる必要がある．

PassStyles は 3 つの課題を解決する．1 つ目の顔画像生成の多様性に関する課題は，ユーザの指定する特徴を含む StyleGAN の層 (キーレイヤ) を特定することで解決する．PassStyles は，データセットから選択された顔画像の潜在ベクトルに，特定されたキーレイヤを挿入することで，数枚の画像から多種多様な認証に用いる画像を生成することができる．次に，3×3 配置によって 2 つ目のショルダサーフィン耐性に関する課題を解決する．3×3 配置とは，非正規ユーザの注意をそらすために，3 つのよく似た顔画像からなる 3 つのグループを認証に用いることである．1 つの正解画像と 2 つのダミー画像からなるグループが 1 つと，3 つのダミー画像からなる 2 つのグループで構成される．各グループ内の画像は互いに類似しているため，正解画像は非正規ユーザの注意を引かないように他のダミー画像に紛れて目立たないように配置される．最後に，3 つ目の誤った画像生成に関する課題は，表示画像調整機能によって解決される．ユーザの認証失敗時にエラーとなる元画像を画像データセットから排除することで認証失敗率を下げることができる．

2. 背景

2.1 グラフィックパスワード

グラフィックパスワードにおいては，ショルダサーフィンに弱いという欠点がある．特にスマートフォンのようなタッチ型の入力インタフェースにおいては，ユーザが画面をタッチした場所から容易にパスワードが漏洩してしまう [1]．ショルダサーフィンに強い認証システムとして，生成図形を用いたワンタイムパスワードを石井の研究 [4] では提案している．しかし，ショルダサーフィンを行った際には，非正規ユーザであっても 27.8%の確率で認証を突破できてしまう．

本論文で提案する PassStyles は，Cognometric 方式であるが，1 節で述べた既存研究 ([9], [10], [11]) のよう正解画像を直接見せることはしない．さらに，PassStyles は認証のたびに正解画像が変化する．そのため，非正規ユーザによるショルダサーフィンが難しくなる．

ユーザビリティと脆弱性への対策を両立させるという目標は，1 回限りの生成図形を用いる石井の研究 [4] と共通している．石井の研究では，認証前にユーザが秘密として形状パターン生成規則 (ルール) を決定する．認証の段階では，システムはルールに合致する正解図形とルールに合致しないダミー図形を生成し，ユーザはルールを満たす正解図形を選択する必要がある．石井の研究では，非正規ユーザの認証率 27.8%（72.2%の防御成功率）を達成したが，生成ルールは 12 種類しか作成することができない．我々の PassStyles は，より多くの種類の条件を定義しつつ，同程度の防御率を達成することを期待する．

2.2 StyleGAN

StyleGAN においては，シード s_A から潜在ベクトル $\boldsymbol{z}_A$ が生成され，潜在ベクトル $\boldsymbol{z}_A$ から中間表現 $\boldsymbol{w}_A$ が生成される．$\boldsymbol{w}_A$ を用いて画像 i_A が生成され，使用するシードが異なると異なる画像が生成される．さらに，2 つの異なる中間表現の両方を用いて画像を生成すると，2 つの画像の特徴の両方を持った画像が生成される．例えば，中間表現 $\boldsymbol{w}_A$ から画像 i_A が生成され，中間表現 $\boldsymbol{w}_B$ から i_B が生成されるとする．このとき，$\boldsymbol{w}_A$ と $\boldsymbol{w}_B$ の両方を画像生成に用いることによって i_A と i_B の特徴をもった画像が生成される (図 1)．2 つの中間表現を混ぜる方法を StyleMixing と呼ぶ.

© 2025 Information Processing Society of Japan.

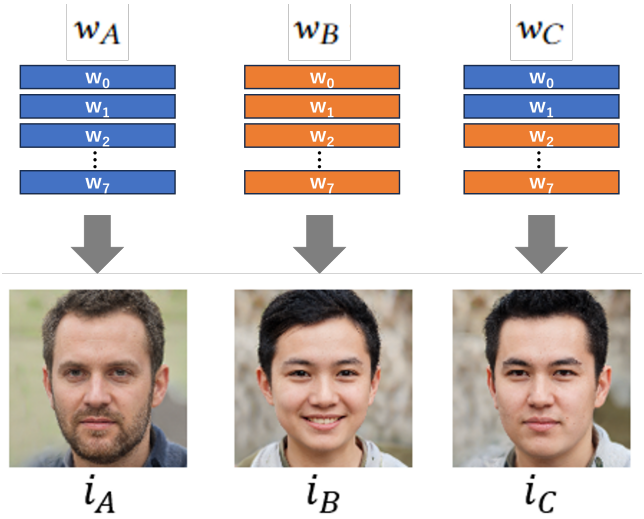

図 1 StyleMixing の例：i_A と i_B の要素を混合させることで i_C を生成する

3. 提案

3.1 概要

PassStyles は 4 つのモジュールから構成される (図. 2).

- パスワード生成モジュール (Generation Module)
- パスワード登録モジュール (Registration Module)
- パスワード認証モジュール (Authentication Module)
- 表示画像調整モジュール (Online Learning Module)

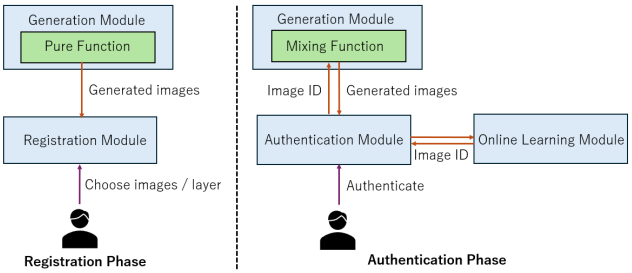

図 2 PassStyles 全体図：左側が登録時，右側が認証時であり，4 種類のモジュールが存在する．

4 つのモジュールについての詳細を説明する．

パスワード生成モジュール: StyleGAN を用いて画像生成を行う．単一生成，混合生成の 2 種類の画像生成方法がある．

- 単一生成: 1 つのシードから 1 枚の画像を生成する．シード s_A から画像 i_A が生成される．生成された画像はパスワード登録モジュールに送信される．
- 混合生成: 2 つの異なるシードから 1 枚の画像を生成する．シード s_A 及び s_B から s_A と s_B の両方の特徴をもつ画像 i_C が生成される．生成画像はパスワード認証モジュールに送信される．

パスワード登録モジュール: このモジュールでは, PassStyles の登録を行う．最初にユーザはパスワード生成モジュールの単一生成にて生成された画像を一通り眺め，顔画像に関する条件を考える．その後，条件を満たす画像と条件を満たさない画像を混合生成で生成された画像群から選択する．どの画像が選択されたかをパスワード認証モジュールに送信する．また，StyleGAN のどの層に特徴が入っているかをユーザは実際の生成画像を見ることで判定でき，どの層を用いるかも認証モジュールに送信する．

パスワード認証モジュール: このモジュールでは，ユーザが認証を実際に実施する. 具体的には，以下の図 3 の画面で認証を行う．画面には 9 枚の顔画像が表示されるので，ユーザは自信の考えた条件に最も合致する画像を選択する．認証モジュールで表示する画像は，登録モジュールから与えられた画像番号，層番号 (キーレイヤ) を元に，生成モジュールの混合生成に画像を与えて，生成モジュールからの出力を元に認証画面を作成する．

図 3 実際の認証画面：「帽子を被っている人」が条件であり，右下画像が正解画像

表示画像調整モジュール: このモジュールでは，認証モジュールにてユーザが誤った画像を選択した場合に，同じ画像群を表示しないようにする機能である．

3.2 事前準備

パスワード登録の前準備として，生成モジュールの単一生成にて，StyleGAN2 [6] を用いて顔画像を N_U 枚生成する．

3.3 パスワード登録

ユーザは，生成モジュールによって生成された 500 枚の画像を一通り眺め (図 4)，顔画像に関する条件を考える．その後，条件を満たす画像を 4 枚，条件を満たさない画像

© 2025 Information Processing Society of Japan.

を 4 枚選択する．

図 4 パスワード登録時に表示する画面

3.4 パスワード認証

本節では，実際にユーザが認証を実施するときに表示される画像の構成について説明する．認証時に表示される画像は 9 枚あり，条件を満たす 1 枚の画像を選択することで認証が進む．ここで，条件を満たす 1 枚の画像を正解画像と定義し，条件を満たさない残りの 8 枚の画像をダミー画像と定義する．9 枚の画像の中から 1 枚の正解画像を 4 連続で選択した場合に認証成功とする．

3.5 キーレイヤの決定

本節では，StyleGAN においてユーザの選択した画像からキーレイヤを決定する方法について説明する．キーレイヤとは，認証に用いる生成画像を生成するときに必要な層と定義する．

生成モジュールの混合生成において，2 枚の画像を固定したまま，使用する層を全ての組み合わせで試したものが図 5 となっている．最も上にある画像が，i_A であり，帽子を被っていない．最も下にある画像が i_B であり，帽子を被っている．すなわち，下にいくほど，i_B の画像に近くなり，右に行くほど，後半の層において i_B の要素が含まれていることになる．

ここで，帽子を被っているという条件を保持したまま i_B と最も類似度が低い画像がどれになるかを考えると，3 行目の最も左にある顔画像となる．これをキーレイヤとする．キーレイヤを固定したまま，i_A を変化させ，帽子を被っている画像であるほかの i_B でも同様のことをしたのが図 6 になる．これを見ると，i_A や i_B によらずキーレイヤが同じであれば，i_B の帽子を被っているという特徴をもつことが分かる．

実際のパスワード登録においては，図 5 において，条件を満たす画像の中で，最も右上にある層を特定しキーレイヤとする．

3.6 3 × 3 配置

非正規ユーザが正解画像を選択するのを難しくする手法として 3 × 3 配置の説明を行う．

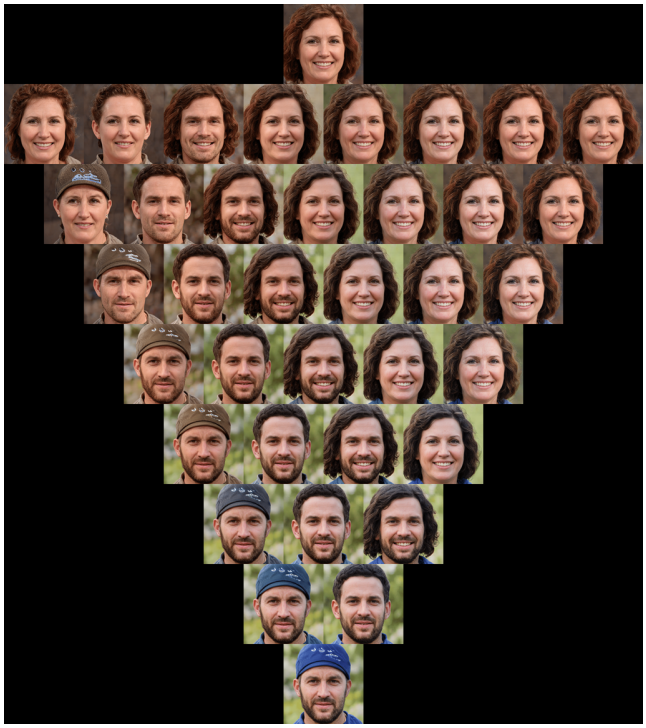

図 5 StyleGAN におけるキーレイヤ決定時に表示する画像

図 6 キーレイヤを固定したときの生成画像

9 枚の画像は，3 種類の類似画像の組 3 つから構成されており，ダミー画像 8 枚のうち 2 枚だけが正解画像に類似している．正解画像に似ているダミー画像と似ていないダミー画像の両方が存在することによって，正解画像がどれか分かりにくくすることが出来る．実際に表示される画像の例は図 3 のようになる．

3.7 オンライン学習

ランダム画像の中に条件を満たすものが含まれていると，ダミー画像の中に条件を満たす画像が生成されてしまうことがあり，正規ユーザの認証成功率が下がる．これを防ぐために，表示画像調整を提案する．表示画像調整においては，ユーザがダミー画像を選んでしまった場合に，ダミー画像に含まれているランダム画像のシードを以降の認証では使わないようにする．

4. 評価

本稿では，人間顔画像データセットを用いて事前学習された，Sony が提供する NNabla (Neural Network Li-

© 2025 Information Processing Society of Japan.

braries) [7] を事前学習モデルとして使用した．事前学習モデルの出力画像の解像度は，1024 × 1024 であるが，メモリ容量節約のため，128 × 128 にダウンサンプリングしたものを実験に使用した．

4.1 条件の提案

4.1.1 提案された条件

19 人の実験参加者 (男性 9 人，女性 10 人) を研究室内で募集した．実験参加者の平均年齢は 22.58 歳 (SD=2.68) である．実験参加者は 3 種類の異なる条件を考え，実際に PassStyles を使用した．

合計で 57 種類の条件が 19 人の実験参加者から生成された．意味が重複している条件を取り除くと 38 種類の条件となる．2 名以上の実験参加者から提案された条件を表 1 に示す．意味がほぼ一致している表現については 1 つの条件にまとめている．

表 1 複数名の実験参加者から提案された条件一覧

ID	条件	人数
1	眼鏡をかけている人	7
2	子供	4
3	髭のある人	3
4	無表情な人	3
5	眼鏡をかけている人	3
6	男性	2
7	短髪の人	2
8	金髪の人	2
9	パーマをかけている人	2

実験参加者が作成したパスワード (表 1) に対して，実験参加者自身が 25 回の画像選択を行ったところ，57 種類のうち 40 種類の条件にて一度も誤りなく正解画像を選択できた．

4.2 ショルダサーフィン耐性実験

4.2.1 概要

本実験では，実際のスマートフォン (iPhone 11) を用いて非正規ユーザが PassStyles の条件を突破することの難易度を測定する．図 7 は実際の実験の様子である．左の人物が正規ユーザであり，右の人物が非正規ユーザである．非正規ユーザは任意の位置から正規ユーザの認証画面を覗き見ることができる．

4.2.2 実験参加者

11 人の実験参加者 (男性 7 人，女性 4 人) が本実験に参加した．実験参加者の平均年齢は 21.27 歳 (SD=0.62) であり，実験参加者は全員初めて PassStyles に触れる．

4.2.3 実験方法

実験参加者は前節の実験で生成された条件のうち，5 種類の条件で実験を行う．最初に，PassStyles を十分慣れている人物が正規認証者として 4 回の画像選択を行う．実験参加者は非正規ユーザとして，正規認証者の操作を自由な位置から覗き見てルールを推測する．正規認証者の画像選択が完了した後に，実験参加者は 8 回の画像選択を行う．実験参加者の画像選択が完了した後，実験参加者は推測した条件をシステムに入力し，正解率と正解の条件が画面に表示される．ここまでを 1 セットとし，顔画像の条件を変更して 5 セットの実験を実施する．各セットで用いる条件は順序効果を防ぐために実験参加者ごとにランダム化している．

図 7 実機実験の様子：左側の人物が正規ユーザであり，右側の人物が非正規ユーザである

4.2.4 実験結果

正解率と認証成功率は表 2 の通りとなった．なお，正規ユーザは認証を一度も間違えなかった．

表 2 スマートフォン実験の結果：非正規ユーザによる正解率と認証成功率．認証成功率は最初の 4 回の画像選択を一度も間違えなかった人の割合を示す．単位は%

条件	正解率	認証成功率
歯が見えている人	45.5	27.3
帽子を被っている人	79.5	72.7
背景が緑の人	14.8	9.1
短髪な人	12.5	0.0
額が前髪で隠れている人	26.1	9.1

表 2 は，PassStyles のショルダサーフィン耐性を実機実験した結果である．値が小さいほど，ショルダサーフィンの耐性があるということを意味し，例えば，条件「短髪な人」では，認証成功率が 0.0%となっており，誰も認証成功出来なかったということを意味している．

5. 考察

石井らの生成図形を用いた研究 [4] では，4 カテゴリ 12 種類の条件しか作ることが出来なかった．一方，PassStyles ではより多くの種類の条件を可能であり，大幅に条件の多様性が向上している．また，今回提案されたパスワード

© 2025 Information Processing Society of Japan.

は 19 人の実験参加者によって提案されたものであり，提案されたパスワード以外の条件も容易に考えられる．例えば．性別に関しては，今回は男性しか提案されなかったものの，女性という条件も容易に考えられる．髪の色に関しては，金髪という条件が提案されたが，黒髪，茶髪，白髪という条件も生成可能である．他にも複数の条件を組み合わせて，眼鏡をかけた男性や金髪でパーマをかけている人といった条件も生成可能であり，PassStyles において登録できるパスワードの多様性は高い．

ショルダサーフィン実験においては，短髪な人，背景が緑の人，額が前髪で隠れている人の 3 条件に関しては，認証成功率が十分に低く，ショルダサーフィン耐性が高い．一方で，葉が見えている人という条件に関しては，27.3%の非正規ユーザが認証を突破できており，帽子を被っている人という条件に関しては，72.7%の非正規ユーザが認証を突破できてしまっている．条件によってショルダサーフィン耐性が大きく異なるため，実際の使用状況においては，パスワード登録時にショルダサーフィン耐性の推定値をユーザに示し，ショルダサーフィン耐性が低い条件であれば，登録をやり直すというシステムを作成する必要がある．ショルダサーフィン耐性を自動判定するシステムの作成は将来研究とする．

6. おわりに

本稿では，ユーザが顔画像に関する条件を考え，9 枚の顔画像から条件を満たす 1 枚の正解画像を繰り返し選択することで認証を行う，直感的で使いやすいグラフィカルパスワードの一種である PassStyles を提案した．PassStyles は，StyleGAN によって生成された多種多様な画像と 3×3 配置により，非正規ユーザが条件を特定することが困難であるため，ショルダサーフィンに強い．パスワードの登録では，ユーザに StyleGAN のどのレイヤに条件が含まれているか，すなわちキーレイヤを特定するインタフェースを提供し，少数の画像から認証用の多種多様な画像を生成することができる．評価実験の結果，様々な画像の特徴を PassStyles の条件に使用できることが示された．また、PassStyles が正規ユーザにとって十分に使用可能であり，非正規ユーザによるショルダサーフィンに対しても頑健であることが確認された．

謝辞 本研究は，JST 次世代研究者挑戦的研究プログラム JPMJSP2123 及び JST，CREST，JPMJCR19A1 の支援を受けたものである．

参考文献

[1] Adam J. Aviv, John T. Davin, Flynn Wolf, and Ravi Kuber. Towards baselines for shoulder surfing on mobile authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference*, ACSAC '17, p. 486–498, New York, NY, USA, 2017. Association for Computing Machinery.

[2] L. Bošnjak, J. Sreš, and B. Brumen. Brute-force and dictionary attack on hashed real-world passwords. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1161–1166, 2018.

[3] Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. Is a picture really worth a thousand words? exploring the feasibility of graphical authentication systems. *International journal of human-computer studies*, Vol. 63, No. 1-2, pp. 128–152, 2005.

[4] Kentaro Ishii. Shoulder-surfing resistant graphical authentication based on one-time shape-pattern generation. *Information Science and Applied Mathematics*, Vol. 30, pp. 13–27, 2023.

[5] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[6] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8110–8119, 2020.

[7] Takuya Narihira, Javier Alonsogarcia, Fabien Cardinaux, Akio Hayakawa, Masato Ishii, Kazunori Iwaki, Thomas Kemp, Yoshiyuki Kobayashi, Lukas Mauch, Akira Nakamura, Yukio Obuchi, Andrew Shin, Kenji Suzuki, Stephen Tiedmann, Stefan Uhlich, Takuya Yashima, and Kazuki Yoshiyama. Neural network libraries: A deep learning framework designed from engineers' perspectives, 2021.

[8] Allan Paivio, Timothy B Rogers, and Padric C Smythe. Why are pictures easier to recall than words? *Psychonomic Science*, Vol. 11, No. 4, pp. 137–138, 1968.

[9] Adrian Perrig Rachna Dhamija. Déjà vu: A user study using images for authentication. *USENIX Security Symposium*, 2000.

[10] Furkan Tari, A Ant Ozok, and Stephen H Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on Usable privacy and security*, pp. 56–66, 2006.

[11] Hideki Koike Tetsuji Takada. Awase-e: Image-based authentication for mobile phones using user's favorite images. *Human-Computer Interaction with Mobile Devices and Services*, pp. 347–351, 2003.

[12] Khoa Trieu and Yi Yang. Artificial intelligence-based password brute force attacks. In *MWAIS 2018 Proceedings*, 2018.

© 2025 Information Processing Society of Japan.