

ゲーム状況に合わせて AI プレイヤに 豊かな表情をとらせるシステム

奥田真¹ 橋本剛¹

概要: ゲームには表情を介した心理戦の面白さがある。こうした表情による面白さを表現するために、AI プレイヤに表情によるリアクションをさせるゲームが存在する。これらのゲームでは、ゲームを動かす AI に連動させて表情を変化させる手法がとられている。しかし、この手法はゲーム AI が無いと成立しない。そこで本論文ではゲーム AI なしに AI プレイヤの表情変化を行うシステムを提案する。人間プレイヤーの顔が映ったゲームプレイ映像を教師データに、ゲーム画面を入力、表情データを出力とした機械学習モデルを事前学習させておき、これをゲーム中に推論することで表情変化を行う。提案システムの構成要素について技術的検討を行い、テトリスを対象に提案システムを実装し、ゲーム AI を使わなくとも、表情変更ができていたことを確認した。

A system that allows AI players to make rich facial expressions in response to game situations

OKUDA MAKOTO¹ HASHIMOTO TSUYOSHI¹

Abstract: Games are interesting in their psychological warfare through facial expressions. To express this interest through facial expressions, there are games that allow AI players to react with facial expressions. These games use a method in which facial expressions are changed in conjunction with the AI that runs the game. However, this method is not feasible without game AI. In this paper, we propose a system that changes the facial expressions of AI players without game AI. A machine-learning model is pre-trained using game-playing videos of human players' faces as teacher data, game screens as input, and facial expression data as output. We conducted a technical study of the components of the proposed system, implemented the proposed system on the game Tetris, and confirmed that the proposed system was able to change facial expressions without using game AI.

1. はじめに

ゲームには表情を介した心理戦の面白さがある。表情による面白さを表現するために、AI プレイヤに表情によるリアクションをさせるゲームが少なからず存在する。ランプの魔神が質問を投げかけてきて、プレイヤーが思い浮かべている人物やキャラクタを絞り込み推測し言い当ててくる Akinator^{*1} では、ランプの魔神がゲーム中に悩む表情

を見せる。6段階ほどの悩み具合の表情イラストが用意されており、プレイヤーが質問に答えるたびイラストを切り替えることでリアクションしている。コンピュータ将棋ソフトの激指^{*2} では、プロ棋士をモチーフにした AI プレイヤとの対戦モードにおいて、対局中に AI プレイヤが明るい表情、暗い表情を見せる。悩んだり喜んだりするプロ棋士の画像がいくつか用意されており、これを切り替えてリアクションをしている。どちらの事例も表情表出と対応するゲームパラメータに連動させて表情画像を切り替えていると予想され、Akinator は推測具合と表情イラストの悩み具

¹ 松江工業高等専門学校
National Institute of Technology, Matsue College

^{*1} UpSwell, Inc.: akinator (online), available from
< <http://www.upswell.jp/service-akinator.php> >
(accessed 2022-09-18).

^{*2} マイナビブックス: 『将棋レポリビューション 激指 15』 (online),
available from < <https://book.mynavi.jp/gekisashi15/> >
(accessed 2022-10-10).

合を、激指は局面評価値と表情の明暗を連動させていると思われる。このように、AI プレイヤにリアクションをさせるためには、ゲームを動かす AI に連動させて表情を変化させる手法がとられている。しかし、この従来手法はゲーム AI が無いゲームで用いることはできず、ゲーム AI そのものから作る必要があり、その難易度は高い。単純に強い AI に着目してみても、これまで将棋や囲碁のような完全情報ゲームでは強いゲーム AI が作られてきたが、不完全情報ゲームはこれからという段階である。それにゲーム AI が作れたとしても、適切な表情表出ができるとは限らない。例えば将棋でプロ棋士並みの実力を持つゲーム AI が使っても、それに連動させた表情は初心者プレイヤに理解できるとは限らない。そこで本論文ではゲーム AI に依存せず AI プレイヤの表情変化を行うシステムを提案する。人間プレイヤの顔が映ったゲームプレイ映像を教師データに、ゲーム画面を入力、表情データを出力とした機械学習モデルを事前学習させておき、これをゲーム中に推論することで表情変化を行う。機械学習モデルにはゲーム画面をそのまま画像として入力する。このアプローチは将棋で例えば、局面から直接表情を生成するようなもので、従来手法のように表情とゲームパラメータを連動させる必要はなくなる。また、提案システムを処理がしやすく、リアルタイム性の高いテトリス^{*3}を対象に実装する。提案システムの構成要素について技術的な検討を行い、具体的な実装内容を述べる。

2章で関連研究について述べ、3章でシステムのアイデアを提案する。その後4章で提案システムの技術的検討を行い、5章でテトリスを対象に実装した提案システム「Tetris2Face」について報告し、6章でまとめを行う。

2. 関連研究

ゲームに限らずいくつかの分野で動的な表情表出の試みが行われている。本章ではなんらかのデータを機械学習モデルに入力し、動的に表情データを生成している研究および事例を紹介する。

大曾根らは表情で親近感を持たせるポーカーパートナーエージェントを提案している [1]。勝率推定の値、ハッタリの可否などの9つのパラメータを入力として、眉、目、口など8種類の表情パーツの位置座標を出力するニューラルネットワークを学習させている。教師データにはアンケート調査の結果を用いており、アンケート対象者はポーカーをプレイし、ターンごとに自分がどのような表情になるか回答する形式をとっている。この研究では、1章で例示した従来手法と同様に、表情と対応するゲームパラメータを選定して機械学習モデルの入力としている。

CG 分野では 3D アバターの表情表現は手間と時間がか

かることで知られており、これを自動化するツール Audio2Face が NVIDIA から提供されている。音声データを事前に学習されたディープニューラルネットワークへ入力することで、表情アニメーションを生成する [2]。Audio2Face は 3D アバターの顔にある、20 個以上のメッシュ頂点を駆動させ表情を形成している。

ロボット分野には人とのコミュニケーションを円滑にするためロボットに表情をとらせる研究がある。ロボットの表情もデザインに手間と時間がかかるため、松井らは人による外部刺激を入力、まぶたや首の回転角などロボットの表情制御値を出力するモデルを事前に学習させておき [3]、これをリアルタイムに推論することでロボットの表情を自動で変更するシステムを提案している。人間の感情は現在の状態だけでなく、過去の状態も影響し決定していると考えられることから、時系列データを取り扱える機械学習モデル Simple Recurrent Network を用いている。入力の外部刺激とは感情のことで、入力された感情に基づいて自然な表情遷移を行うことを目的としたシステムを実装している。システムでは、モニターに表示された赤、青、シアン、緑色のパネルをクリックすることで、それぞれの色に対応した怒り、悲しみ、驚き、喜びの4感情が入力され、表情遷移がおこる。

いずれの研究でもニューラルネットワークの構造を含む機械学習モデルが使われており、表情データの生成にはニューラルネットワークが適していることを示唆している。

3. システムのアイデア提案

AI プレイヤに表情によるリアクションをさせるゲームでは、ゲームを動かす AI に連動させて表情を変化させる手法がとられている。しかしこの従来手法の適用可否はゲーム AI の有無に依存しており、これを解決するためゲーム AI を作ろうとしてもその難易度は高く、適切な表情表出ができるとは限らない。

そこでゲーム AI に依存せずに AI プレイヤの表情変化を行うシステムを提案する。提案システムの構成図を図 1 に示す。人はゲーム中に表情をとるとき、多くのゲームパラメータを複合的にみて表情を決定していると考えられる。そこでゲームを遊ぶ人をひとつの表情決定システムと捉え、これをゲーム画面を入力、表情データを出力とした機械学習モデルで実現する。人間プレイヤの顔が映ったゲームプレイ映像を教師データに、機械学習モデルを事前学習させておき、これをゲーム中に推論することで表情変化を行う。機械学習モデルにはゲーム画面をそのまま画像として入力する。このアプローチは将棋で例えば、局面から直接表情を生成するようなもので、従来手法のように表情とゲームパラメータと連動させる必要はないため、ゲーム AI なしに表情変更ができる。

^{*3} Tetris Holding: Tetris (online), available from < <https://tetris.com/> > (accessed 2022-10-08).

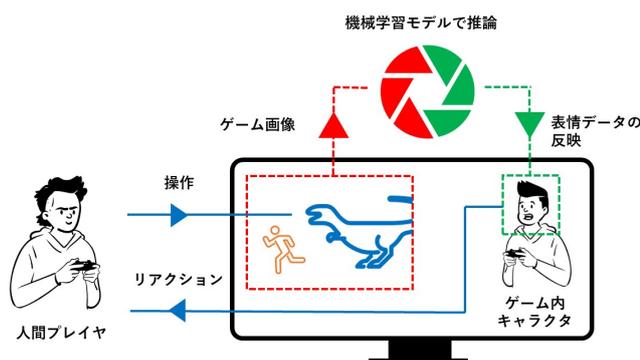


図 1 自動表情生成システム概要図

Fig. 1 Schematic diagram of automatic facial expression generation system

4. 技術的検討

本章ではデータの取り扱いや学習モデル、教師データの集め方など、技術的検討が必要な提案システムの構成要素について述べる。

4.1 表情表現

本章では AI プレイヤの見た目である 3D モデルの表情表現の方式について述べる。機械学習モデルが出力する表情データは、

- 顔パーツの動きを数値で示す方式
- 感情の度合いを示した値で表情を表現する方式

の 2 種類の方式のどちらか、もしくは両方を用いて表現できる。本論文では 1 つ目の出力方式を「顔パーツ方式」、2 つ目を「感情値方式」と呼ぶものとする。

どちらの方式も Ekman らが提案した仕組みや考えに基づくものであり、顔パーツ方式は顔面動作符号化システム FACS (Facial Action Coding System) [4] に、感情値方式は基本 6 感情 [5] の考えに基づいている。FACS は人の顔に現れる表情を解剖学的視点から分析し、AU (Action Unit) と呼ばれる顔面筋の動作有無で表情を符号化する仕組みで、複数の AU を組み合わせることであらゆる表情を表現できる。AU には AU1, AU2, AU3, といった番号が 46 番まで割り振られている。例えば AU4 は「眉を下げる」、AU44 は「薄目にする」ことに対応しており、AU4 と AU44 によりにらむ表情を表現できる。基本 6 感情とは「喜び」「悲しみ」「怒り」「嫌悪」「恐怖」「驚き」の 6 種類の感情のことであり、Ekman らは「人類みな基本 6 感情に対応した表情をとれる」という人類共通の普遍則を提案している。この普遍則を受け入れるならば、基本 6 感情という共通のプロトコルでどんな人の表情もある程度表現可能だといえるため、FACS, AU とともに、表情デザインをする多くのソフトウェアで基本 6 感情を取り扱う方式が導入されている。

Unity ^{*4}, UnrealEngine ^{*5} といったポピュラーなゲームエンジンでも、顔パーツ方式と感情値方式の両方で表情表現ができるようになっている。

顔パーツ方式は、AU の部位と動作に対応した顔パーツを動かし、表情を表現する。図 2 に「右目をウインクする」のパラメータを 0.0 から 1.0 まで 0.25 刻みで変化させた場合の例を示す。パラメータと右目まぶたの閉じ具合が連動しており、0.0 の全開状態から、1.0 の全閉状態の間の遷移を表現できる。2 章でとりあげたポーカーの研究 [1]、音声データによる表情表現の事例 [2] はこの方式を採用している。

感情値方式はあらかじめ感情に対応した表情をプリセットとして顔パーツ方式でデザインしておき、感情の度合いを 0.0 から 1.0 の範囲で指定することで表情を表現する。図 3 に「喜び」のパラメータを 0.0 から 1.0 まで 0.25 刻みで変化させた場合の例を示す。0.0 の真顔から、1.0 の笑顔までの遷移を表現できるが、その遷移の間の表情は目や口が中途半端に開いた状態となるため、思う通りの喜び具合を表現できない場合がある。2 章でとりあげたロボットの研究 [3] はこの方式を採用している。

2 つの方式のメリット、デメリットについて述べる。顔パーツ方式は顔パーツの細かい動きを表現可能で自由度は高い。その分制御すべきパラメータが多くなり、人力の表情デザインでは手間と時間がかかる。また、パラメータが多いのでモデルの学習が難しくなり、推論速度も下がる機械学習面でのデメリットも考えられる。一方で感情値方式はとらえたい表情の感情さえわかれば、比較的簡単に目的の表情を取らせることができる。複数種の感情の混交表現も可能で、例えば泣き笑いなら喜びと悲しみのパラメータを両方高くすればよい。しかし、図 3 で例示したように、微かな感情の機微や、顔パーツの細かい動きを表現するのは難しい。このように自由度は低いとその分機械学習モデルが出力するパラメータは少ないので、学習が比較的簡単で、推論速度も速いという機械学習面でのメリットがある。

4.2 機械学習モデル

物理パラメータと感性パラメータの関係を学習させるには、ニューラルネットワークが適しているとされており [6]、ゲーム画面と表情データという質の異なるデータ間の関係を学習させる場合もニューラルネットワークが有望であると考えられる。2 章で取り上げた研究と事例でもすべてニューラルネットワークの構造を含んだモデルを用いており、その有望さを示唆している。また松井らは、人間

^{*4} Unity Technologies: Unity (online), available from < <https://unity.com/ja> > (accessed 2022-10-08).

^{*5} Epic Games, Inc.: Unreal Engine (online), available from < <https://www.unrealengine.com/ja/> > (accessed 2022-10-08).

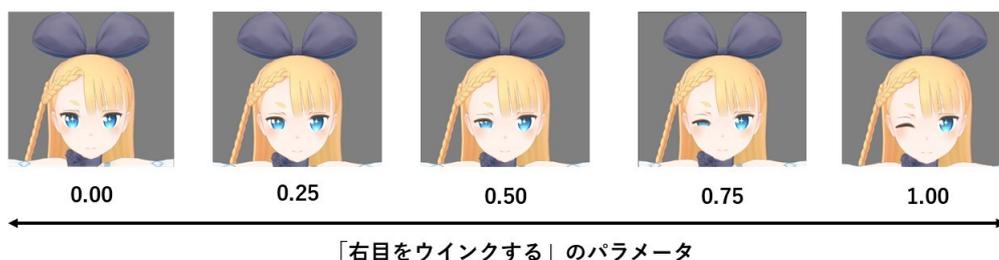


図 2 顔パーツ方式：「右目をウインクする」のパラメータを変化させた例

Fig. 2 Face part method: Example of changing the parameter for "wink right eye"

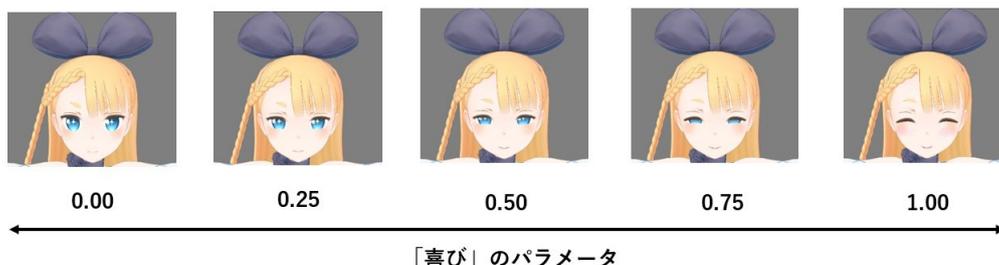


図 3 感情値方式：「喜び」のパラメータを変化させた例

Fig. 3 Emotion value method: Example of changing the parameter for "joy"

の感情は現在の状態だけでなく、過去の状態も影響し決定していると考えられることから、時系列データを処理可能な Simple Recurrent Network を採用している [3]。このことから、ニューラルネットワークから検討を始め、Simple Recurrent Network などニューラルネットワークの発展形のモデルへと検討を進めていくことが妥当だと考えられる。

4.3 教師データ

教師データには人間のゲームプレイ映像を用いるが、映像には表情とゲーム画面が安定して明瞭に映り続けることが肝要である。データの収集にはインターネット上のゲーム実況プレイ動画、ゲーム大会動画を用いる方法と、自分でゲームプレイ映像を撮影する方法の 2 種類が考えられる。前者はインターネット上で近年大幅増加しているため、データの大量確保が比較的簡単にできることが期待できるが、編集によって表情とゲーム画面が明瞭に映らない場合があるのがネックとなる。後者は人力であるためデータの大量確保に比較的時間がかかるが、自分の処理のしやすいようにデータを準備可能である。

5. 提案システム実装：Tetris2Face

本論文では提案システムをテトリスを対象に実装した。これを本論文では「Tetris2Face」と呼ぶものとする。テトリスは以下 3 点の理由から採用した。

- リアルタイム性が高く、表情の切り替わりも高頻度であるため。
- 単純でデータ処理しやすいため。

- 後述する良質な教師データが大量に得られる見込みがあったため。

リアルタイムで状況が変わるゲームは表情も頻繁に変わることが予想されるため、提案システムの題材として適している。しかし、FPS ゲームや格闘ゲームなどに代表されるリアルタイム性の高いゲームは、複雑でデータ処理しにくいことが多い。その点テトリスはリアルタイム制の高さと単純さを両立する、本研究に適したゲームのひとつといえる。本章では Tetris2Face の実装について述べる。

5.1 機械学習モデル：T2FNet

機械学習モデルにはニューラルネットワークを採用し、これを「T2FNet」と呼ぶものとする。教師データにはテトリスの世界大会、CTWC (Classic Tetris World Championship) *6 の動画を用いる。CTWC の動画は以下 2 点の理由から教師データとして採用した。

- 図 4 に示すように、表情とゲーム画面が安定して映り続けるため。
- インターネット上に複数動画がアップロードされているので、今後データの大量確保が見込めるため。

今回学習対象にするのは左側の選手のみで、30 分の動画 1 本分を 1 秒ごとに画像で書き出し、1800 枚の画像を用意した。

教師データの具体的な処理と、その事前学習のイメージを図 5 に示す。入出力データには学習を容易にするための

*6 CTWC: Classic Tetris World Championship (online), available from < <https://thectwc.com/> > (accessed 2022-10-08).

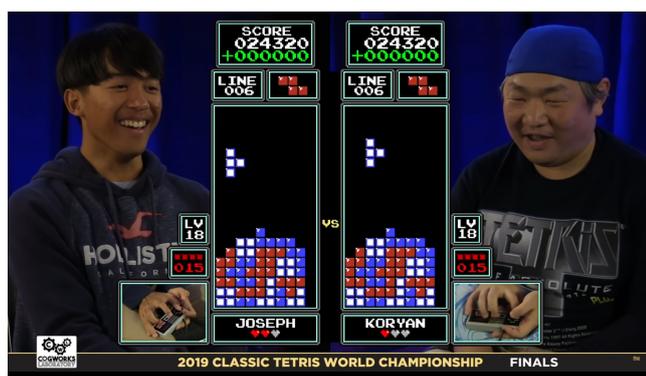


図 4 教師データ：テトリスの世界大会「CTWC」

Fig. 4 Training data: CTWC, the world Tetris tournament

工夫を行っている。入力ของเกม画面は画像処理により盤面のみを入力するようにし、盤面はパターンマッチングによりテトリミノの有無を1,0で表現した one-hot 表現に変換している。テトリスの盤面は横10マス、縦20マスであるため、 $10 \times 20 = 200$ の長さのビット列を T2FNet に入力する。出力の表情データは感情値方式で、感情は6感情のうち「喜び」のみを採用し、*7Py-feat *8 という感情認識ライブラリを用いて顔画像を感情値に変換している。このように図4からゲーム画面部と表情部の情報を抽出し、T2FNet の事前学習を行う。

T2FNet のニューラルネットワークは、層数は5層、各層におけるユニット数はそれぞれ200,1120,1568,352,1で、学習係数は0.01、活性化関数はすべて RELU 関数とした。ここまでのハイパーパラメータは keras-tuner *9 というライブラリを用いて Hyperband によるチューニングを行い決定した。ネットワークの学習は誤差逆伝搬法で2000エポック行い、損失は平均絶対誤差で0.23になった。

5.2 Tetris2Face の構成と実装

本章ではテトリスゲーム、AI プレイヤの見た目である3D アバター、そして推論エンジンからなる表情生成システム、Tetris2Face の具体的な実装方法について述べる。システム構成を図6に示す。ポピュラーなゲームエンジンである Unity を採用し、テトリスの盤面データの取得から推論までをすべてゲーム内で完結するように実装している。事前準備として、事前学習した T2FNet を ONNX (Open Neural Network eXchange) *10 という機械学習モデルの

*7 今回収集した教師データの選手がとる表情の感情は「喜び」が主であったため。

*8 Jolly, E., Cheong, J. H. and Xie, T.: Py-Feat: Python Facial Expression Analysis Toolbox - Py-Feat (online), available from < <https://py-feat.org/pages/intro.html> > (accessed 2022-10-08).

*9 Keras: KerasTuner (online), available from < https://keras.io/keras_tuner/ > (accessed 2022-10-08).

*10 The Linux Foundation: Open Neural Network Exchange (online), available from < <https://onnx.ai/> > (accessed 2022-10-08).

汎用フォーマットでエクスポートし、Unity 上で動作する推論エンジン Unity Barracuda *11 で読み込んでおく。また、AI プレイヤの見た目である3D アバターは VRM *12 という3D アバターの汎用フォーマットで読み込み用意しておく。*13ゲームを起動するとテトリスゲームがスタートし、テトリスの盤面を描画するスクリプトから盤面データを one-hot 表現で抜き出して、これを T2FNet に逐次的に入力する。そして推論を行い、喜びの感情値を出力する。その後感情値は3Dモデルの顔に反映され、感情値方式で表情を表出する。こうして人間プレイヤーはテトリスをプレイすることで、キャラクターの表情によるリアクションを見ることができる。

5.3 感情表出の確認

Tetris2Face の動作の様子を図7に示す。画像左側のグラフは T2FNet の出力を示しており、横軸に時間[s]、縦軸に0.0から1.0までの「喜び」の感情値を示している。ゲームAIを使わなくとも、表情変更ができていることを確認した。

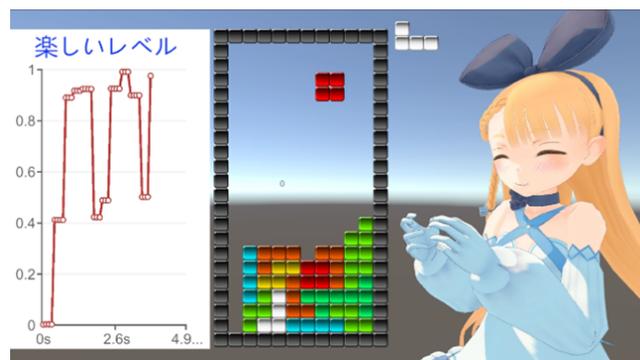


図 7 Tetris2Face が動作する様子

Fig. 7 Tetris2Face in action

6. まとめ

本論文ではゲーム AI なしに AI プレイヤの表情変化を行うシステムを提案した。技術的検討を行ったのち、テトリスを対象に提案システム「Tetris2Face」を実装し、ゲーム AI を使わなくとも、表情変更ができていることを確認した。今後はニューラルネットワークの発展形の機械学習モデルでの学習など、より自然な表情表現を目指すための実装を進めていく。

*11 Unity Technologies: Introduction to Barracuda (online), available from < <https://docs.unity3d.com/Packages/com.unity.barracuda@1.0> > (accessed 2022-10-08).

*12 VRM Consortium: VRM ドキュメント (online), available from < <https://vrm.dev/> > (accessed 2022-10-08).

*13 VRM はあらゆるプラットフォームで利用できるように設計されており、汎用性に長けるため採用した。

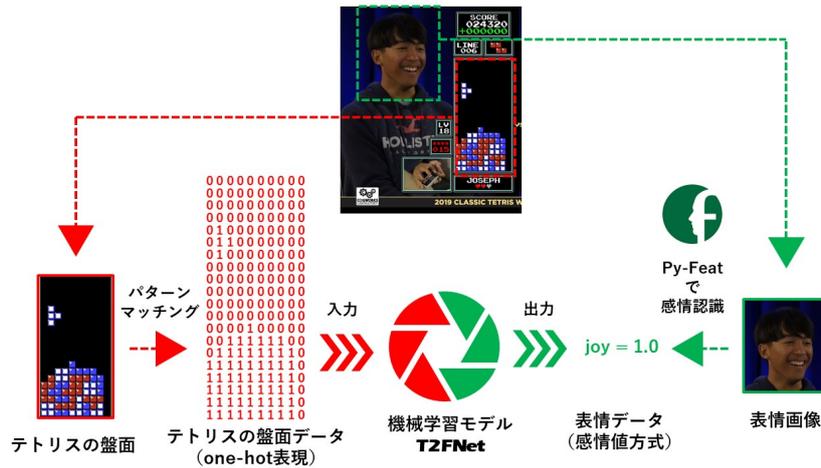


図 5 教師データの処理

Fig. 5 Processing of training data

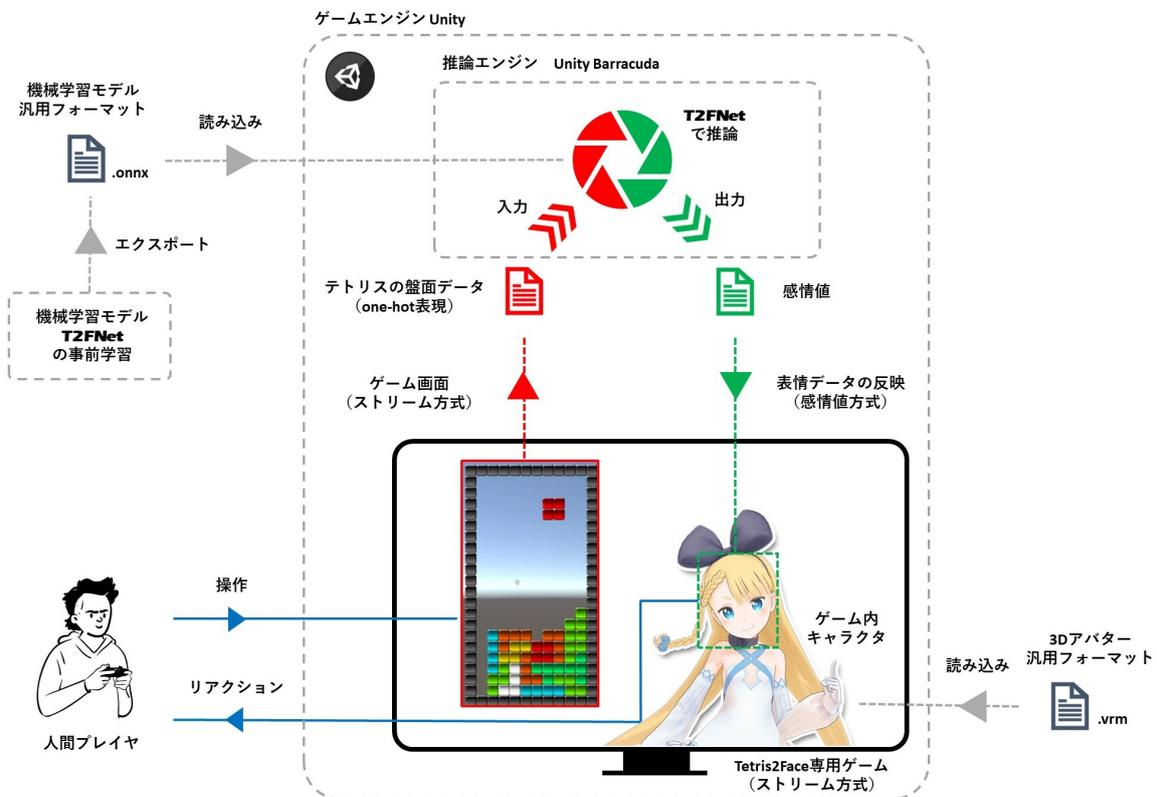


図 6 Tetris2Face システム構成

Fig. 6 Tetris2Face System Configuration

参考文献

[1] 大曾根圭輔, 鬼沢武久: ユーザに親近感を持たせるボーカールパートナーエージェント, 日本知能情報ファジィ学会 ファジィシステムシンポジウム 講演論文集, Vol.21, No.6, pp.1127 - 1142, 2009.

[2] NVIDIA: Omniverse Audio2Face (online), available from < <https://www.nvidia.com/ja-jp/omniverse/apps/audio2face/> > (accessed 2022-10-06).

[3] 松井裕紀, 加納政芳, 加藤昇平ほか: Simple Recurrent Network を用いた感性ロボットのインタラクティブ表情表出, 日本ロボット学会誌, Vol.28, No.3, pp.360 - 368, 2010.

[4] Ekman, P., and Friesen, W. V. : Facial Action Coding System (FACS), A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press.(2002).

[5] Ekman, P. : facial expressions, Handbook of Cognition and Emotion, pp.301-320 (1999).

[6] 鈴木健嗣, 山田寛, 橋本周司: ニューラルネットワークによる顔面表情の物理的パラメータと感性的パラメータの対応付け, 電子情報通信学会技術研究報告, Vol.HCS2000-47, pp.7-13, 2001.