

AI解析に適した映像データ高圧縮技術 ～クラウド・5Gでの利用に向けて～

久保田 智規 中尾 鷹詔 吉田 英司 久保田 真

富士通が独自に開発した本技術の概要については、2020年3月5日付けプレスリリース「[AIが認識できる画質で映像データを高圧縮する技術を開発](#)」や、FUJITSU JOURNAL「[AI解析用の映像データを従来の10分の1に圧縮！クラウド活用で防犯等の社会課題の解決に貢献](#)」でも既に紹介している。本稿では、技術の部分に焦点を当て、その詳細を説明する。

あらまし

本稿では、従来の映像圧縮技術の問題点とAI解析向け映像データ高圧縮技術の詳細、およびその評価結果について述べる。

近年、街頭での不審者監視や製造ラインでの作業者の行動分析など、様々な現場で映像の利用が増えている。更に、高速・大容量の特性を持つ5Gサービスの開始によって、超高精細な4K映像や多数のカメラ映像が利用可能になることから、これらの映像をAIで解析し、従来、ヒトの目では見落としていた細かな行動も認識可能とすることが期待されている。映像データのAI解析は、多大な計算量を必要とし、またカメラ数の増減によって必要な計算量は変化する。このような計算には、費用が利用分だけで済み、かつスケーラビリティに優れたクラウドが適している。しかし従来は、大きなサイズの映像データをクラウドに送信するとネットワーク帯域を占有してしまうため、クラウドを利用することが難しかった。この課題を解決すべく、富士通独自のAI解析に適した映像データを高圧縮する技術を開発した。本技術の適用によって、クラウドでの映像解析が可能になる。

1. まえがき

近年、街頭や製造ラインなどの様々な現場でカメラ映像の利用が増えている。特に、2020年に国内でも第5世代移動通信システム（5G）[1]のサービスが開始され、高速大容量通信が可能になったことで、今後は4K超高精細映像の利用も爆発的に増えることが見込まれる。また、映像データをAIで解析し、得られる知見をマーケティングや製造業の品質検査など、様々な分野で活用することが活発化すると見込まれる[2-4]。

大量の映像を人手によらず自動的に解析するAI手法としては、深層学習（Deep Learning）が一般的である。深層学習は計算量が多く、また例えばカメラを更に数多く配置して監視の範囲を広げる、あるいは時間帯によって監視箇所を増減させるなど、カメラの数の増減によっても必要な計算量は変化する。このような計算には、費用が利用分だけで済み、かつスケラビリティに優れたクラウドが適している[5]。しかし従来は、大きなサイズの映像データを現場からクラウドに送信すると、ネットワーク帯域を占有してしまうため、クラウドを利用することが難しかった。

こうした問題を解決するために、AIが認識できる画質に最適化することで、これまでのヒトが目視することを前提とした圧縮よりも格段に高圧縮できる富士通独自の技術を開発した[6, 7]。これによって、ネットワーク帯域の占有を軽減し、クラウドの活用が可能となる。

本稿の構成は以下である。まず、従来の映像圧縮技術とその問題点について説明する。次に、AIでの解析に適した映像を高圧縮する技術とその評価結果を示す。最後に、まとめと今後の予定について述べる。

2. 従来の映像圧縮技術とその問題点

本章では、従来の映像圧縮技術と、AIでの解析に適用する際の問題点について述べる。

2.1 従来の技術はヒト向けの圧縮

映像圧縮技術とは、高画質な映像をやり取りする通信コストや、保存するストレージコストの削減などのために、映像を圧縮してデータサイズを小さくする技術である。一般的に、圧縮率が高いほどデータサイズは小さくなる一方、画質は悪くなるため、両者のバランスを取ることが重要となる。

従来は、H.265/HEVC [8]などの映像圧縮方式に関する国際標準規格・勧告（以下、規格）を用いて、「ヒト向けの画質」で圧縮率を決定していた。具体的には、映像の1コマ1コマである画像を格子状に区切った「ブロック」ごとに、ヒトが見て違和感のない画質となる圧縮率を設定する[9]。

2.2 従来の技術はデータサイズの大きさが問題点

図-1は、画像認識などの映像解析用AIの一例であるYOLOv3 [10]が、ある画像の中で人物の位置を認識・検出（以下、認識）する際に注目した画素を、可視化手法 [11]を用いて白い点で表したものである。この結果によると、AIが認識する際に注目する画素は、画像のごく一部であることが分かる。つまり、従来の圧縮技術ではAIによる映像認識には不要なブロックも高画質としているため、映像全体のデータサイズが大きくなってしまっていることが問題であった。

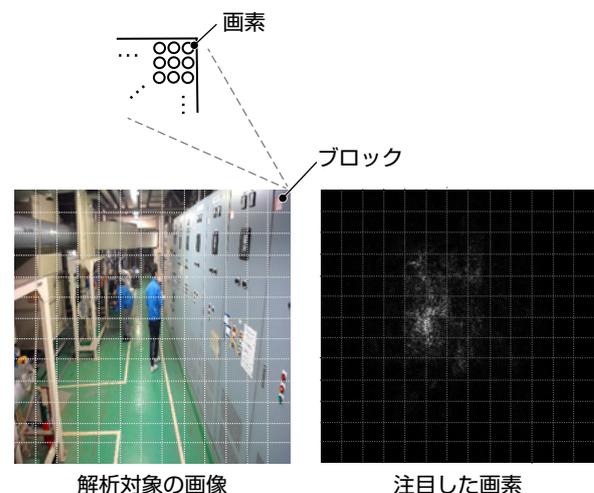


図-1 YOLOv3が注目した画素とブロックの関係

3. AI解析に適した映像データ高圧縮技術

本章では、AI解析に適した画質で映像データを高圧縮する技術について説明する。

3.1 本技術のアプローチ

本技術は二つのステップで実現される。

まず、画像の圧縮率を上げながらAIによる物体認識を試行し、認識確度がどの程度劣化するかを把握する。認識確度とは、AI自身が認識結果をどの程度確かとしたかを示す値で、AIの出力である。本ステップの詳細は3.2節で説明する。

次に、非圧縮時と同様にAIが物体を認識できる範囲で最も高い圧縮率（限界圧縮率）を決定する。データサイズを極力小さくできるように、この圧縮率はブロック単位できめ細かく決定する。最後に、

決定した限界圧縮率で各ブロックを圧縮する。本ステップの詳細は3.3節で説明する。

3.2 圧縮率とAIの認識確度劣化の関係を把握

本節では、圧縮率と認識確度劣化の関係を把握する三つの処理について、図-2を用いて説明する。

(1) 圧縮による認識確度の劣化度合いの数値化

まず、圧縮していない画像（非圧縮画像）全体をある一様な圧縮率で圧縮し、その後復元する。復元後の画像を圧縮後復元画像と呼ぶ。圧縮時には情報が失われるため、圧縮後復元画像の画質は、圧縮前より劣化している。次に、非圧縮画像と圧縮後復元画像のそれぞれを入力としてYOLOv3などの深層学習モデルによって物体を認識する処理を行い、認識結果および認識確度を得る。最後に、非圧縮画像と圧縮後復元画像の認識確度の差を求め、非圧縮画像からの認識確度の劣化度合いを数値化する。

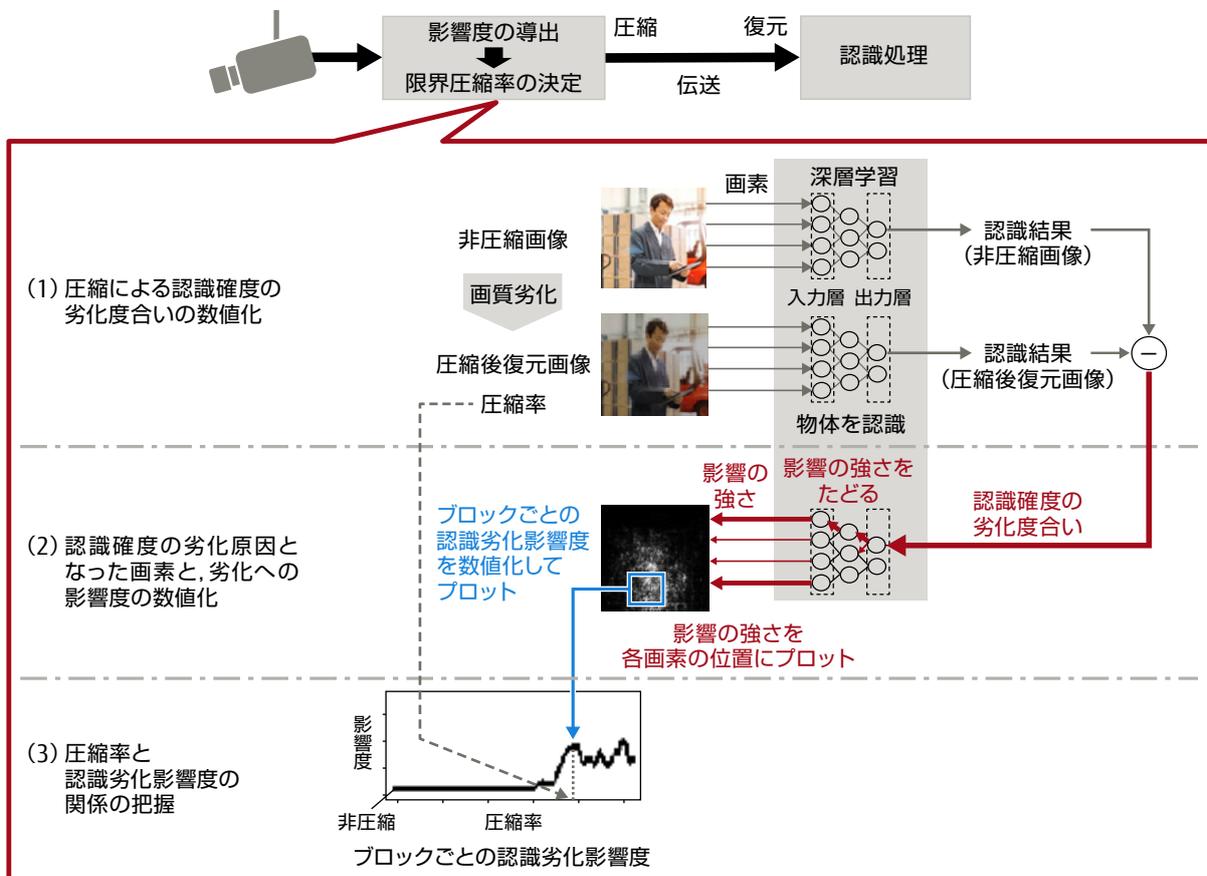


図-2 圧縮率とAIの認識への影響の把握

深層学習モデルでは、認識対象として学習したパターン（例えば、人物、車など）全てに対して、入力した画像がどの程度の確度で当てはまるかの確率値（0～1）を出力する。これが認識確度であり、認識確度が最も高いパターンを認識結果とする。そして、非圧縮画像と圧縮後復元画像から求めた認識確度の差分値が、非圧縮画像からの劣化度合いとなる。

(2) 認識確度の劣化原因となった画素と、劣化への影響度の数値化

まず、認識確度を劣化させる原因となった画素、およびその画素の認識確度劣化への影響を、誤差逆伝播法 [12] を用いて調べる。具体的には、深層学習モデルの出力層に (1) で求めた認識確度の劣化度合いを入力し、物体を認識する処理の流れを逆方向にたどる。これによって、認識確度の劣化の原因となった画素の位置と、その画素が認識確度の劣化

に与えた影響の強さ（以下、認識劣化影響度）を得ることができる。次に、画素ごとの認識劣化影響度を、ブロック単位で集計する。これによって、各ブロックの認識劣化影響度を数値化できる。

(3) 圧縮率と認識劣化影響度の関係の把握

(1) ～ (2) で行った認識劣化影響度の数値化を、同じ画像に対して圧縮率を変化させながら実行する。これによって、各ブロックの圧縮率と認識劣化影響度の関係を図-2に示すグラフのように把握することができる。

3.3 限界圧縮率の決定と画像の圧縮

本節では、3.2節で求めた圧縮率と認識劣化影響度の関係を用いて、物体を認識できる限界圧縮率をブロックごとに決定し、各ブロックを圧縮する処理について説明する。限界圧縮率の決定、および画像の圧縮の全体像を図-3に示す。

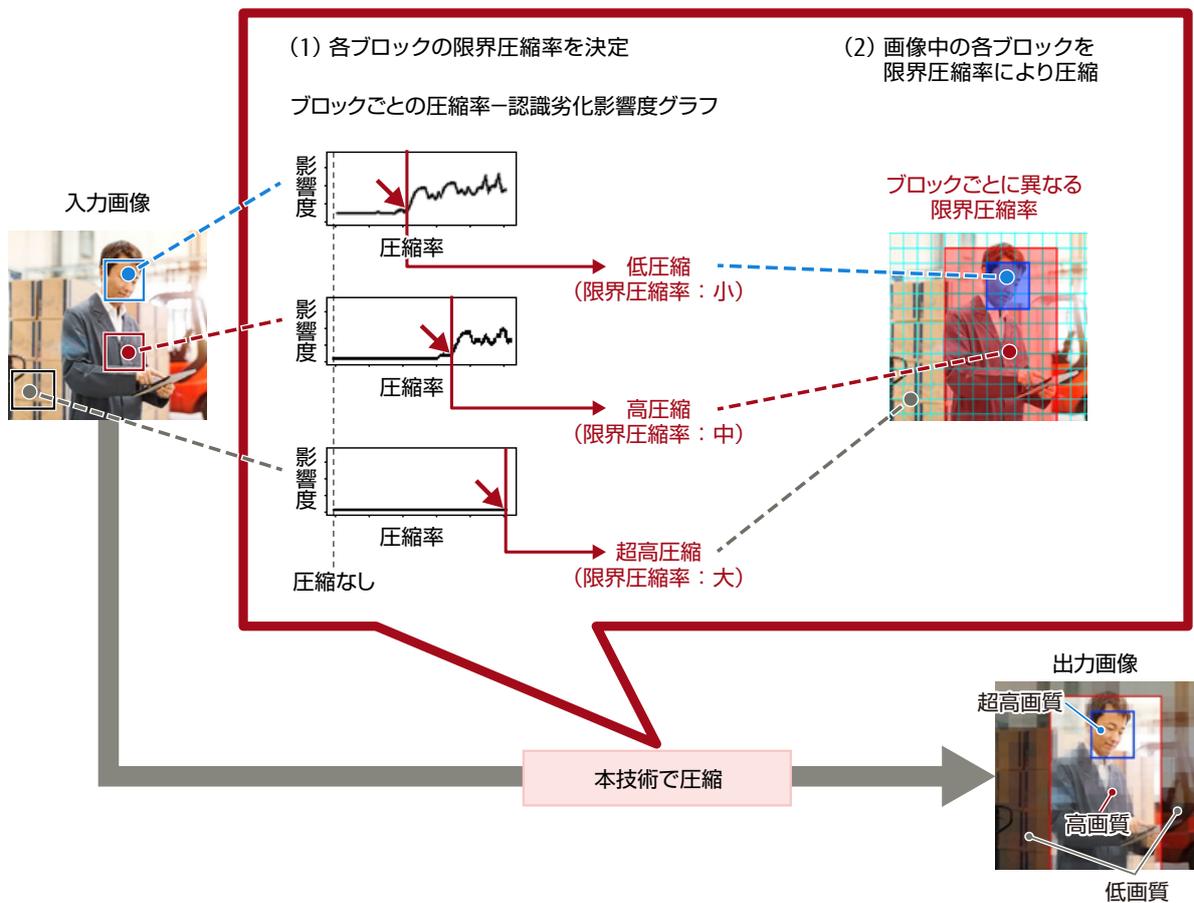


図-3 限界圧縮率の決定と画像の圧縮

(1) 各ブロックの限界圧縮率を決定

3.2節で求めたグラフにおいて、圧縮率を変化させた際に、ある圧縮率で認識劣化影響度が急増する場合、以下の理由によって、限界圧縮率をそれ以下にする。

- ・各ブロックの認識劣化影響度は、非圧縮の場合が最小である
- ・認識劣化影響度が大きなブロックほど、認識に必要な情報が多く含まれ、圧縮によってこの情報が失われやすいことを意味する
- ・ある圧縮率で認識劣化影響度が急増する場合、認識確度も急激に劣化するため、この圧縮率を超えてはならない

以上の基準を適用して限界圧縮率を決める処理を、図-3の人物とその顔を認識する例で説明する。各ブロックの圧縮率－認識劣化影響度グラフにおいて、圧縮率を低から高に変化させる。この場合、まず人物の顔近辺のブロックで認識劣化影響度が急増し、次に顔以外の人物のブロックで急増する。これらのブロックにおける限界圧縮率は、それぞれで急増した圧縮率以下とする。一方で、背景のブロックでは圧縮率を高くしても認識劣化影響度は変わらないため、限界圧縮率は指定可能な最大値とする。

(2) 画像中の各ブロックを限界圧縮率で圧縮

(1)で求めた限界圧縮率を、圧縮率を設定する既存の規格に従い、画像の各ブロックに設定して圧縮する。

3.2節～3.3節の処理は、画像1コマごとに行う。そのため、時間と共に人物が移動する場合でも、コマごとの人物の位置に応じてリアルタイムに、各ブロックの限界圧縮率を変えることができる。

以上の技術によって、認識確度を維持したまま映像データの高圧縮が実現できる。

4. 評価

本章では、今回開発した技術による映像データの圧縮率と認識結果の評価について説明する。

4.1 データサイズと認識結果を評価

本評価では、YOLOv3で切り出した画像中の人物から、CPN (Cascaded Pyramid Network) [13]

という深層学習モデルで骨格を認識し、2種類の評価を行った。まず、従来のヒトが監視する際に一般的に適用する圧縮率と、本技術で決定した圧縮率のそれぞれで映像を圧縮し、圧縮後のデータサイズを比較した。次に、圧縮後復元画像を骨格認識処理に入力し、認識結果を比較した。

4.2 データサイズは1/7でも認識結果は同じ結果に

ヒト向けの圧縮データのサイズが11,814バイトであるのに対して、本技術では1,527バイトで、データサイズを約1/7に削減できていることが確認できた。

図-4は、従来のヒト向けの圧縮 (a) と本技術を用いた圧縮 (b) のそれぞれについて、圧縮後復元画像に骨格認識処理が出力した骨格情報を重畳した画像の例である。本評価では、本技術でデータサイズをヒト向けの圧縮の約1/7に削減しても、映像の全てのコマにおいて、人物の骨格を非圧縮の映像とほぼ同じ形状で認識できた。

4.3 考察

本節では、映像中のある1コマを抽出した画像 (図-5) を使って、データサイズを約1/7に削減できたポイントを説明する。

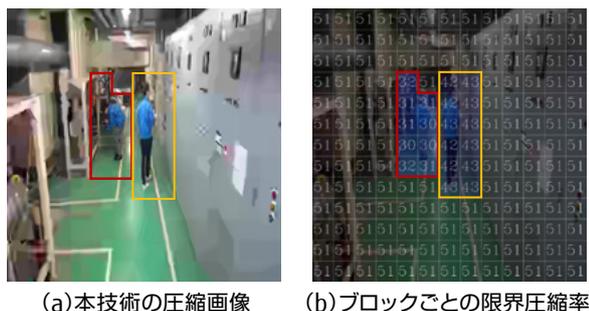
図-5 (a) は本技術による圧縮画像、図-5 (b) は、本技術によって求めたブロックごとの限界圧縮率である。圧縮率はH.265/HEVCで設定可能な範囲である0～51を用い、大きな値ほど圧縮率が高い。

図-5 (b) から、骨格認識に影響の大きなブロックだけが30～43と低い値に設定され、他は全て



(a)ヒトによる監視用途向けの圧縮 (b)本技術による圧縮

図-4 ヒト向けと本技術による骨格認識の結果



(a)本技術の圧縮画像 (b)ブロックごとの限界圧縮率
注) 赤枠/橙色枠は本技術によりAIが認識する際に重要なブロックと判断した領域を示す。

図-5 本技術による圧縮処理の考察

最も高い51に設定されていることが分かる。30～43のばらつきは、3.3節で説明した方法で、本技術が認識劣化影響度の違いを用いて圧縮率をきめ細かく判断した結果に基づく。一方、比較対象としたヒト向けの圧縮画像では、全ブロックの圧縮率が30に設定されていた。

したがって、AI認識において重要な部分は画像中の一部の領域であり、その他の重要でない領域は超高圧縮としたこと、更に重要な部分もきめ細かく圧縮率を変えたことで、従来に比べて圧縮率を高くすることが可能になったと言える。

ただし、圧縮の効果は画像サイズ、認識対象の数・大きさ、画像の動きの複雑さなどによって変わる可能性がある点は留意されたい。

5. まとめと今後の予定

本稿では、AI認識にとって重要な部分とそうでない部分を検出し、圧縮率をきめ細かに制御することで、AIが認識できる画質へ最適化し、映像データを高圧縮する技術について述べた。

本技術を用いることで、大きなサイズの映像データを現場からクラウドに送信することが可能となり、クラウド上でのAI解析が可能になる。また、解析結果をセンサーデータ、文書情報などクラウド上の他の情報、例えば映像解析による店舗内の顧客行動と売り上げ情報を組み合わせてマーケティングに活用するなど、より高度な解析・活用が可能となる。

更に、5Gが普及することで、例えば工場のライン変更時にケーブルレス化した多数のカメラの設置場

所を容易に変更できるなどの利点が生じ、本技術による多数のカメラの映像データを高圧縮する効果と合わせて、超高精細映像の利用・活用の敷居を下げることが可能となる [14]。

今後は、本技術を様々なユースケースに適用すべく、AIの認識対象の撮影条件、大きさ、動きなど、実適用で想定される各種条件で評価し、更なる圧縮性能向上の研究開発を進めていく。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献・注記

- [1] 高速大容量・多数同時接続・低遅延を特長とする移動通信システム。
- [2] 総務省：平成30年版 情報通信白書 (2018).
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/pdf/index.html>
- [3] 科学技術振興機構 研究開発戦略センター：研究開発の俯瞰報告書 システム・情報科学技術分野 (2019年).
<https://www.jst.go.jp/crds/pdf/2018/FR/CRDS-FY2018-FR-02.pdf>
- [4] 萩原直彦：第5世代移動通信システム (5G) の今と将来展望。総務省 (2019)。
- [5] 経済産業省：平成29年度我が国におけるデータ駆動型社会における基盤整備 (情報処理実態調査の分析及び調査設計等事業) 調査報告書 (2018).
https://www.meti.go.jp/statistics/zyo/zyouhou/result-2/pdf/H29_report.pdf
- [6] 富士通：AIが認識できる画質で映像データを高圧縮する技術を開発 (2020).
<https://pr.fujitsu.com/jp/news/2020/03/5.html>
- [7] FUJITSU JOURNAL：AI解析用の映像データを従来の10分の1に圧縮！クラウド活用で防犯等の社会課題の解決に貢献 (2020).
<https://blog.global.fujitsu.com/jp/2020-06-04/01/>
- [8] ISO/IEC 23008-2:2017, High efficiency coding and media delivery in heterogeneous environments— Part 2: High Efficiency Video Coding. Oct. 2017. | Recommendation ITU-T H.265 (2018), High Efficiency Video Coding (2018).
- [9] 山田悦久：映像符号化評価用標準動画像. 映像情報メディア学会誌, Vol. 62, No. 8, pp. 1283-1285 (2008).

https://www.jstage.jst.go.jp/article/itej/62/8/62_8_1283/_pdf/-char/ja

[10] J. Redmon et al. : YOLOv3: An Incremental Improvement. CoRR, Vol. abs/1804.02767.

<https://arxiv.org/pdf/1804.02767.pdf>

[11] K. Simonyan et al. : Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. CoRR, Vol. 1312.6034 (International Conference on Learning Representations (ICLR) 2014 Workshop) (2013).

<https://arxiv.org/pdf/1312.6034.pdf>

[12] David E. Rumelhart et al. : Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press (1986).

[13] Y. Chen et al. : Cascaded Pyramid Network for Multi-Person Pose Estimation. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7103-7112 (2018).

<https://arxiv.org/pdf/1711.07319.pdf>

[14] 富士通：ローカル5G (2020).

<https://www.fujitsu.com/jp/innovation/5g/local5g/>



吉田 英司 (よしだ えいじ)

株式会社富士通研究所
ICTシステム研究所
データシステム技術の研究・開発に従事。



久保田 真 (くぼた まこと)

株式会社富士通研究所
デジタル革新コア・ユニット
エッジ・クラウドを活用したデジタルサービスの研究・開発に従事。

この記事は、富士通の技術情報メディア「富士通テクニカルレビュー」に掲載されたものです。他の記事も是非ご覧ください。

富士通テクニカルレビュー

<https://www.fujitsu.com/jp/technicalreview/>



著者紹介



久保田 智規 (くぼた ともりの)

株式会社富士通研究所
デジタル革新コア・ユニット
リアルタイムメディア処理システム技術、AI応用システム技術、AIのふるまい解析技術の研究・開発に従事。



中尾 鷹詔 (なかお たかのり)

株式会社富士通研究所
デジタル革新コア・ユニット
映像システム技術、AI応用システム技術、ストレージシステム技術の研究・開発に従事。