AI 特許紹介(55)

AI 特許を学ぶ！究める！

～拡散モデルを用いた SR3 特許～

2023 年 8 月 10 日

河野特許事務所

所長弁理士　河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第４次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許出願人　Google

出願日　2023 年 1 月 17 日

公開日　2023 年 5 月 18 日

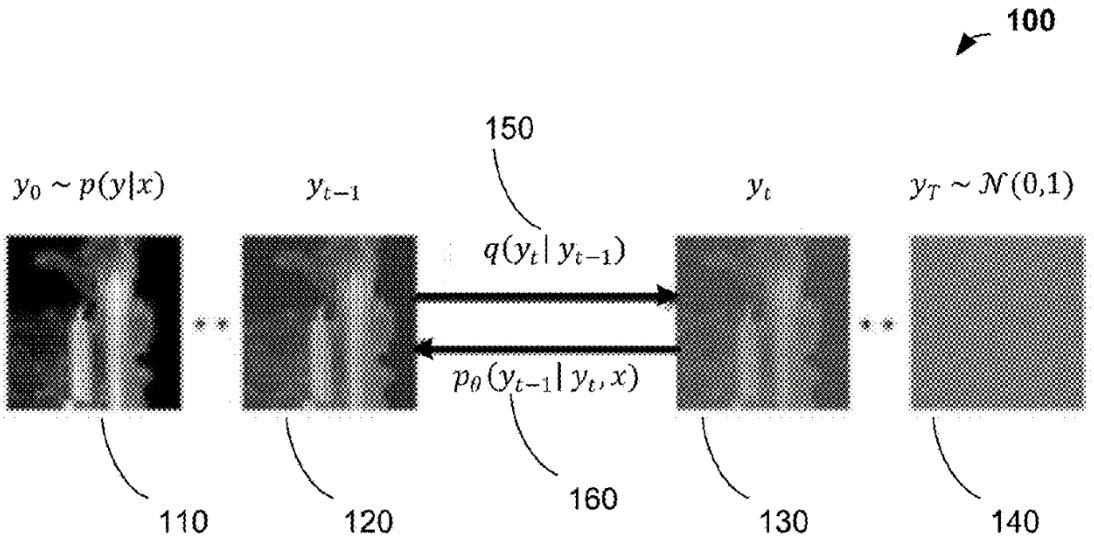
公開番号 US20230153959

発明の名称　機械学習モデルに基づく反復改良による画像強化

959 特許は、ノイズ除去拡散確率モデルを条件付き画像生成に適応させ、確率的反復ノイズ除去プロセスを通じて超解像度を実行する SR3(Super-Resolution via Repeated Refinement)技術に関する。

2.特許内容の説明

図１は、順拡散プロセスおよび反復ノイズ除去プロセスの一例を示す図である。



入出力画像ペアの特定のデータセットは、次のように示され、未知の条件付き分布 $p(y|x)$ から抽出されたサンプルを表す。

$\mathcal{D}=\{x_i, y_i\}_{i=1}^{N}$

$p(y|x)$へのパラメトリック近似は、ソース画像 x をターゲット画像 $y\in\mathbb{R}^d$ にマッピングする確率的反復改良プロセスを通じて学習され、ノイズ除去拡散確率（DDPM：denoising diffusion probabilistic）モデルを条件付き画像生成に適応させるアプローチが利用される。

条件付き DDPM モデルは、画像 110 によって表されるターゲット画像 y_0 をＴ個のリファインメントステップで生成する。たとえば、画像 140 で表される純粋なノイズ画像 $y_T\sim\mathcal{N}(0,1)$から開始して、モデルは、画像 130,120 等で示されるように、$p_\theta(y_{t-1}|y_t, x)$として与えられ $y_0\sim p(y|x)$となる学習された条件付き遷移分布 160 に従って、反復$(y_{T-1}, y_{T-2}, \ldots, y_0)$を通じて画像を繰り返し改良する。

入力画像の反復ノイズ除去を可能にするために、トレーニングデータ内の複数の画像ペアのそれぞれの少なくとも１つの対応するターゲットバージョンにガウスノイズを追加する順ガウス拡散プロセスが適用される。例えば、推論チェーンにおける中間画像の分布は、$q(y_t|y_{t-1})$で示される固定マルコフチェーン 150 を介して信号にガウスノイズを徐々に加える前方拡散プロセスの観点から定義することができる。

入力画像の反復ノイズ除去を実行して、入力画像の強化されたバージョンを予測することができる。反復ノイズ除去は、順方向ガウス拡散プロセスに関連付けられた逆マル

コフ連鎖に基づく。

ガウス拡散プロセスは、x を条件とした逆マルコフ連鎖を通じてノイズから信号を反復的に回復することによって逆転することができる。リバースチェーンは、ソース画像とノイズの多いターゲット画像を入力として受け取り、ノイズを推定できるニューラルノイズ除去モデル f_θ を使用して学習できる。

(1)ガウス拡散プロセス

拡散モデルは、ガウスノイズを段階的に追加し、純粋なノイズになるまでデータの詳細をゆっくりと削除し、その後、そのような破損プロセスを逆転させるためにニューラルネットワークをトレーニングすることにより、トレーニングデータを破損するように構成されている。

この反転破損プロセスを実行すると、クリーンなサンプルが生成されるまで徐々にノイズが除去され、純粋なノイズからデータが合成される。この合成手順は、データ密度の勾配に従って、可能性の高いサンプルを生成する最適化アルゴリズムとして解釈できる。T 回の反復にわたって高解像度画像 y_0 にガウスノイズを徐々に追加する順マルコフ拡散プロセス q を定義する。

$$q(y_{1:T} \mid y_0) = \prod_{t=1}^{T} q(y_t \mid y_{t-1}) \qquad \text{(Eqn. 1)}$$

$$q(y_t \mid y_{t-1}) = \mathcal{N}\left((y_t \mid \sqrt{\alpha_t}\, y_{t-1}, (1-\alpha_t)I\right), \qquad \text{(Eqn. 2)}$$

スカラーパラメーター$\alpha_{1:T}$は、$0<\alpha_t<1$ の条件を満たすハイパーパラメータであり、各反復で追加されるノイズの分散を決定する。y_{t-1} は、確率変数の分散が $t\to\infty$として制限されたままとなるように、$\sqrt{\alpha_t}$ によって減衰される。たとえば、y_{t-1} の分散が 1 の場合、y_t の分散も 1 になる。y_0 が与えられた場合の y_t の分布は、次のように中間ステップを周辺化することによって特徴付けることができる。

$$q(y_t|y_0)= \mathcal{N}\left((y_t|\sqrt{\gamma_t}y_0,(1-\gamma_t)I\right), \qquad \text{(Eqn. 3)}$$

$\gamma_t=\Pi_{i=1}{}^{t}\,\alpha_i$

さらに、代数的操作と二乗を完了することにより、(y_0, y_t) が与えられた場合の y_{t-1} の事後分布は次のように導出される。

$$q(y_{t-1} \mid y_0, y_t) = \mathcal{N}\left(y_{t-1} \mid \mu, \sigma^2 I\right) \quad \text{(Eqn. 4)}$$

$$\mu = \frac{\sqrt{\gamma_{t-1}}\,(1-\alpha_t)}{1-\gamma_t} y_0 + \frac{\sqrt{\alpha_t}\,(1-\gamma_{t-1})}{1-\gamma_t} y_t \quad \text{(Eqn. 5)}$$

$$\sigma^2 = \frac{(1-\gamma_{t-1})(1-\alpha_t)}{1-\gamma_t} \quad \text{(Eqn. 6)}$$

この事後分布は、リバースチェーンをパラメータ化し、リバースチェーンの対数尤度の変分下限を定式化するときに有利になる。ニューラルネットワークは、このガウス拡散プロセスを逆にすることを学習する。

(2)ノイズ除去モデルの最適化

拡散プロセスの逆転を可能にするために、ソース画像 x の形式で追加情報を利用し、ソース画像 x とノイズのあるターゲット画像 y を入力として受け取り、ノイズのないターゲット画像 y_0 を回復することを目的としたニューラルノイズ除去モデル f_θ を最適化する。

$$\tilde{y} = \sqrt{\gamma} y_0 + \sqrt{1-\gamma}\,\epsilon, \epsilon \sim \mathcal{N}(0,1), \quad \text{(Eqn. 7)}$$

ノイズを含むターゲット画像 y の定義は、式 3 の順方向拡散プロセスのさまざまなステップでのノイズを含む画像の周辺分布と互換性がある。順方向ガウス拡散プロセスの適用には、反復ステップでのガウスノイズの分散を示すスカラーハイパーパラメータを決定することが含まれる。例えば、ソース画像 x とノイズのあるターゲット画像 y に加えて、ノイズ除去モデル $f_\theta(x, \tilde{y}, \gamma)$ は、ノイズ γ の分散に対する十分な統計を入力として取得することができる。

入力画像の反復ノイズ除去は、順方向ガウスプロセス中に追加されたガウスノイズの分散に基づいてノイズベクトルを予測することを含む。ノイズ除去モデル $f_\theta(x, \tilde{y}, \gamma)$ を訓練して、ノイズベクトル ϵ を予測する。ノイズ除去モデルには、スカラー γ の条件付けを通じてノイズのレベルの情報が提供される。f_θ をトレーニングするために提案された目的関数は次のように記述できる。

$$\mathbb{E}_{(x,y)} \mathbb{E}_{\epsilon,\gamma} \| f_\theta(x, \tilde{y}, \gamma) - \epsilon \|_p^p \quad \text{(Eqn. 8)}$$

$\epsilon \sim \mathcal{N}(0, 1)$, (x, y) は、トレーニング データセット、$p \in \{1, 2\}$、および $\gamma \sim p(\gamma)$からサンプリングできる。γ の分布は、モデルと生成される出力画像の品質に大きな影響を与える。式 8 のように f_θ の出力を ϵ に回帰する代わりに、f_θ の出力を y_0 に回帰することもできる。γ および y が与えられると、ϵ および y_0 の値は相互に決定論的に導出できる。

(3)反復改良による推論

モデルに基づく推論は、順拡散プロセスの逆方向に進む逆マルコフプロセスとして定義することができる。このモデルは、純粋なノイズのみが残るまで（順ガウス拡散プロセスを介して）高解像度画像にノイズが徐々に追加される画像破損プロセスでトレーニングされ、次にモデルはこのプロセスを逆に学習し、純粋なノイズから開始して徐々にノイズを除去し、入力された低解像度画像のガイダンスを通じて目標の分布に到達する。ガウスノイズ y_T から開始すると、次のことが得られる。

$$p_\theta(y_{0:T} \mid x) = p(y_T) \prod_{t=1}^{T} p_\theta(y_{t-1} \mid y_t, x) \qquad \text{(Eqn. 9)}$$

$$p(y_T) = \mathcal{N}(y_T \mid 0, I), \qquad \text{(Eqn. 10)}$$

$$p_\theta(y_{t-1} \mid y_t, x) = \mathcal{N}\left(y_{t-1} \mid \mu_\theta(x, y_t, \gamma_t), \sigma_t^2 I\right) \qquad \text{(Eqn. 11)}$$

推論プロセスは、学習可能な等方性ガウス条件付き分布 $p_\theta(y_{t-1} \mid y_t, x)$に関して定義できる。例えば、ハイパーパラメータを $\alpha_{1:T} \approx 1$ となるように選択することによって、順方向プロセスステップのノイズ分散が可能な限り小さく設定される場合、最適な逆方向プロセス $p_\theta(y_{t-1} \mid y_t, x)$ はほぼガウスになる。

したがって、式 11 で表される推論プロセスでガウス条件式を選択すると、真の逆プロセスに合理的に適合する。一方、$1-\gamma_t$ は十分に大きいため、y_T が式 10 の事前分布 $P(y_T)=\mathcal{N}(y_T \mid 0, I)$に従って近似的に分布し、サンプリングプロセスを純粋なガウスノイズで開始できるようになる。ノイズ除去モデル f_θ は、y_t を含む任意のノイズのある画像 y が与えられた場合に、ノイズベクトル ϵ を推定するように訓練される。したがって、y_t が与えられると、y_0 は式 7 の項を並べ替えることによって近似できる:

$$\hat{y}_0 = \frac{1}{\sqrt{\gamma_t}}\left(y_t - \sqrt{1-\gamma_t}\, f_\theta(x, y_t, \gamma_t)\right), \qquad \text{(Eqn. 12)}$$

推定された $\hat{y}_0$ を式 4 の $q(y_{t-1} \mid y_0, y_t)$ の事後分布に代入して、$p_\theta(y_{t-1} \mid y_t, x)$ の平均を次のようにパラメータ化できる。

$$\mu_\theta(x, y_t, \gamma_t) = \frac{1}{\sqrt{\alpha_t}}\left(y_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_\theta(x, y_t, \gamma_t)\right), \qquad \text{(Eqn. 13)}$$

$p_\theta(y_{t-1} \mid y_t, x)$ の分散は、$(1-\alpha_t)$ に設定でき、デフォルトは、順方向プロセスの分散によって与えられる。このパラメータ化に続いて、SR3 モデルの下での反復改良の各反復は次の形式を取る。

$$y_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}\left(y_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_\theta(x, y_t, y_t)\right) + \sqrt{1-\alpha_t}\,\epsilon_t, \qquad \text{(Eqn. 14)}$$

$\epsilon_t \sim \mathcal{N}(0, I)$.

これは、データ対数密度の勾配の推定値を提供する f_θ を使用したランジュバン力学の 1 ステップに類似する。式 11 で概説した確率モデルに対する式 8 のトレーニング目標の選択は、変分下限の観点とノイズ除去スコアマッチングの観点に基づいて行うことができる。

下記図は、入出力画像 200 の例を示す。入力画像 210 については、SR3 モデルに基づく出力画像 220 が示されている。例えば、入力画像 210 は 16×16 の解像度を有する画像であるのに対し、出力画像 220 は 256×256 の超解像度である。参考画像 230 も示されている。

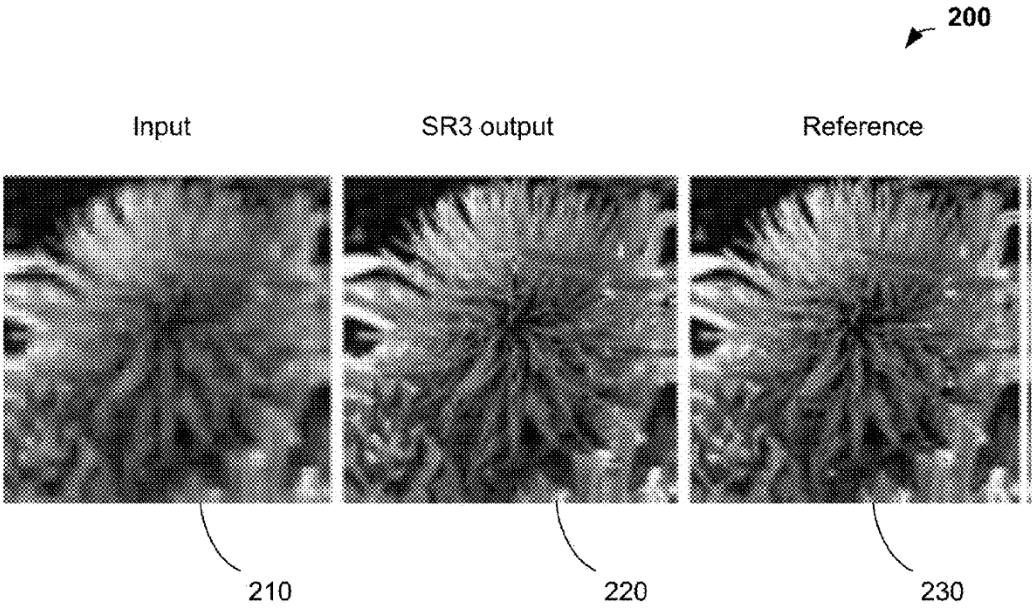


図 3A は、ニューラルネットワークの例示的なアーキテクチャ 300A を示す図である。

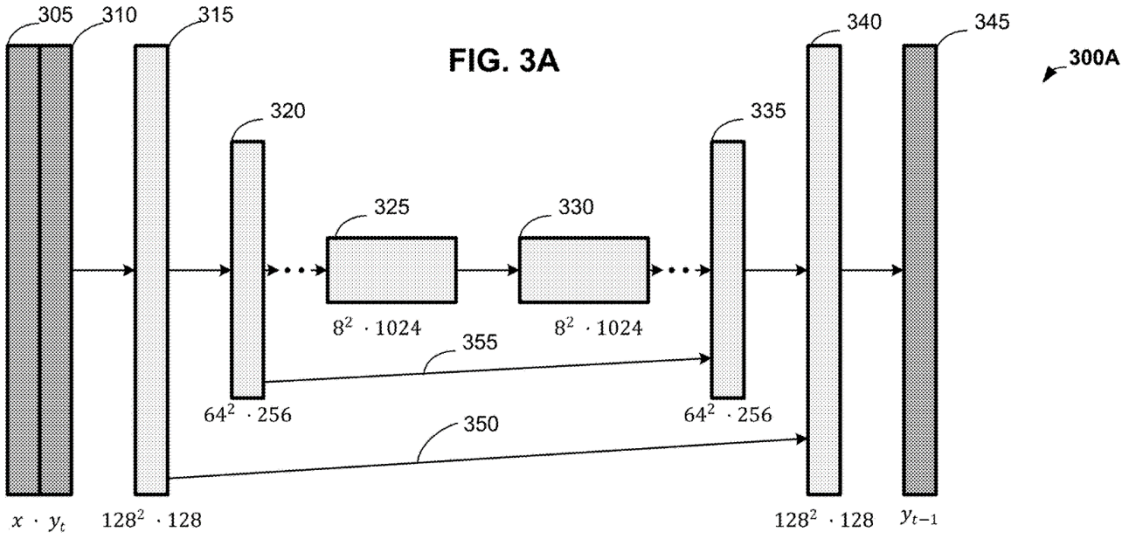


図 3A は、スキップ接続を備えた U-Net アーキテクチャ 300A の説明である。低解像度入力画像 305、x は、目標高解像度に補間され、ノイズを含む高解像度画像 310、yt と連結される。 16×16→128×128 超解像度のタスク例のアクティベーション次元が示される。

ニューラルネットワークは、ノイズ除去拡散確率（DDPM）モデルに基づく U-Net アーキテクチャを備える畳み込みニューラルネットワークであってもよい。例えば、SR3 アーキテクチャは、例えば、DDPM で利用される U-Net などの U-Net300A に基づくことができ、元の DDPM 残差ブロックは、BigGAN からの残差ブロックで置き換えることができ、スキップ接続は、$1/\sqrt{2}$ で再スケーリングされる。

入力 x でモデルを調整するには、バイキュービック補間を使用して、低解像度画像をターゲット解像度にアップサンプリングする。結果は、チャネル次元に沿って yt と結合される。図に示すように、第 1 のノイズを含む高解像度画像 **310**, y_t から第 2 のノイズを含む高解像度画像 **345**, y_{t-1} までの反復の一ステップにおいて、低解像度入力画像 **305**, x は、ブロック 315 で 128×128 からブロック 320 で 64×64 に、ブロック 325 で 8×8 にダウンサンプリングされる。

次に、ダウンサンプリングプロセスからの出力は、ブロック 330 で 8×8 から、ブロック 335 で 64×64 に、そしてブロック 340 で 128×128 にアップサンプリングされる。ブロック 315 をブロック 340 に接続するスキップ接続 350、およびブロック 320 をブロック 335 に接続するスキップ接続 355 など、スキップ接続を使用する。

$$p(\gamma) = \Sigma_{t=1}^{T} \frac{1}{T} U(\gamma_{t-1}, \gamma_t).$$

トレーニング中、タイム ステップ t~{0, . . . , T}を一様にサンプリングし、続いてγ~U(y_{t-1}, γ_t)をサンプリングする。

3.クレーム

959 特許のクレーム 1 は以下の通りである。

1. コンピュータ実装方法において、

画像データベースからトレーニングデータを受信し、

トレーニングデータに基づいて、低解像度の入力画像の高解像度バージョンを予測するニューラルネットワークをトレーニングし、トレーニングには、バイキュービック補間を使用した低解像度入力画像のダウンサンプリングが含まれ、ニューラルネットワークは拡散プロセスに基づいてトレーニングされ、

ノイズ内容が事前定義された閾値を超えるまで、高解像度画像に繰り返しノイズを追加する画像破損プロセスと、

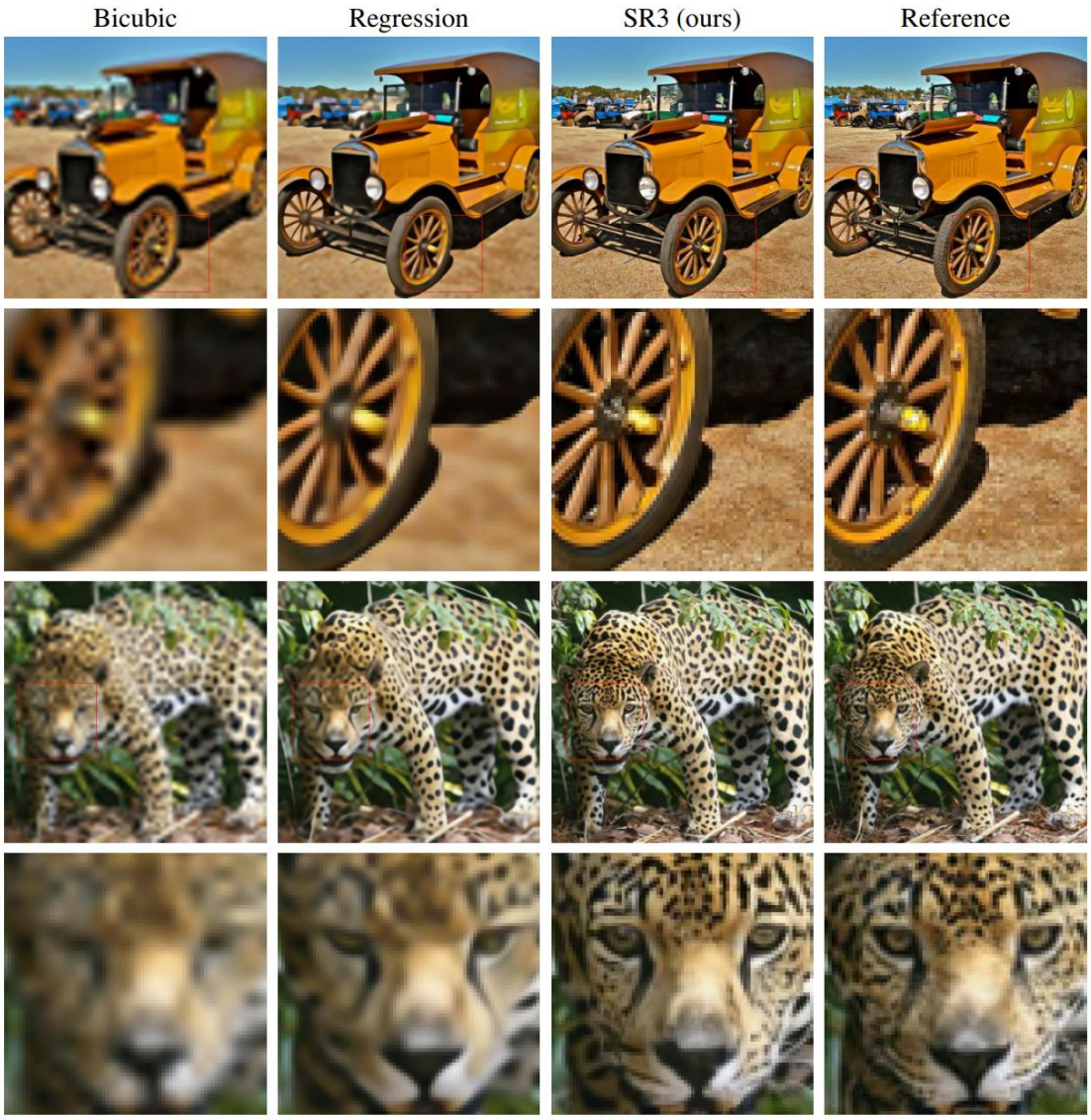
事前定義された閾値を超える初期ノイズ内容を持つ初期画像から開始し、初期画像から繰り返しノイズを除去して目標の分布を達成することによって、画像破損プロセスを逆転することを学習する画像ノイズ除去プロセスとを含み

トレーニングされたニューラルネットワークを出力する。

4. 本特許に関連する論文

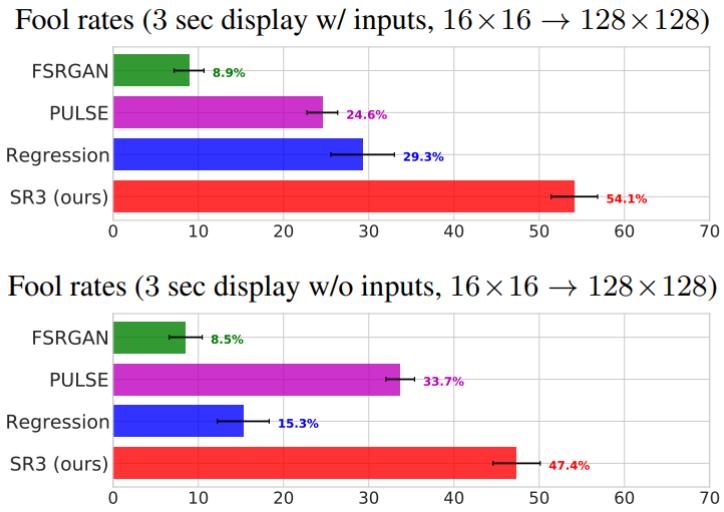
本特許に関する論文“Image Super-Resolution via Iterative Refinement”[1]が、Chitwan Saharia 氏らにより公表されている。下記写真は ImageNet でトレーニングされ、2 つの ImageNet テスト画像で評価された SR3 モデル（64×64 → 256×256）の結果を示す。

[1] Chitwan Saharia et al. “Image Super-Resolution via Iterative Refinement” arXiv:2104.07636v2 [eess.IV] 30 Jun 2021



SR3 は、ノイズ除去拡散確率モデルを条件付き画像生成に適応させ、確率的反復ノイズ除去プロセスを通じて超解像度を実行する。出力の生成は純粋なガウスノイズから始まり、さまざまなノイズレベルでのノイズ除去についてトレーニングされた U-Net モデルを使用してノイズの多い出力を繰り返し調整する。

SR3 は、顔や自然画像など、さまざまな倍率での超解像度タスクで強力なパフォーマンスを発揮する。CelebA-HQ 上の標準的な 8 倍の顔超解像タスクについて人間による評価を実施し、SOTA GAN 手法と比較した。SR3 は下記グラフに示すように 50% に近いフール率を達成し、写真のようにリアルな出力を示唆しているが、GAN のフール率は 34%を超えない。



以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース修了。

AI 特許コンサルティング、医療 AI 特許コンサルティングの他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウエア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門 3」、「ブロックチェーン 3.0(共著)」がある。