The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

不確実性を考慮した深層教師なし異常部分検知

Deep Unsupervised Anomaly Segmentation via Aleatoric Uncertainty-Aware Score

佐藤 一輝 Kazuki Sato　濱 健太 Kenta Hama　松原 崇 Takashi Matsubara　上原 邦昭 Kuniaki Uehara

神戸大学 大学院システム情報学研究科
Graduate School of System Infomatics, Kobe University

Image-based anomaly segmentation is a basic topic in the field of image analysis. Especially, unsupervised training is preferred when dealing with unknown types of anomalies. For this purpose, probabilistic models trained to maximize the likelihood of known samples are employed to identify samples with low estimated likelihoods as anomalies. However, they tend to be sensitive to complex structures rather than semantic anomalies. In this paper, we propose a novel uncertainty-aware score for anomaly segmentation by removing the term that reflects the data complexity from the approximated log-likelihood. Experimental results demonstrate the robustness of the proposed score with respect to data complexity.

1. はじめに

画像解析において，異常部分のセグメンテーションは基本的なテーマの一つである．医用画像解析においては，例えば脳に存在する腫瘍や脳卒中後の病変のような異常部分のセグメンテーションにより，精密な診断や腫瘍成長速度の評価を行うことができる [Menze 15, Bakas 17, Havaei 17]．教師あり学習は画像セグメンテーションのタスクに非常に良い結果を出している [Kendall 17] が，学習にはピクセルごとにラベル付けされた多数のデータを要する上に，学習に用いたデータに存在しないような系統の異常は検出することが困難である．従って，教師データなしで異常部分を検知する必要性が生じる．

サンプル単位での教師なし異常検知では，機械学習により正常なサンプルの分布をモデル化し，その分布から外れたサンプルを異常として検出する．例えば，ある画像に含まれるピクセル $\{x_1, x_2, \dots\}$ の確率モデル $p_\theta(x_1, x_2, \dots)$ は，与えられるサンプル $x \in \mathcal{X}$ がこのモデルから生成される尤度を最大化するよう学習する．確率モデル $p_\theta(x)$ は既知のサンプルに似たサンプルに対して高い尤度を割り当てるため，尤度 $p_\theta(x)$ の低い画像を異常として検出を行う．最近の研究では，深層ニューラルネットワーク (deep neural network; DNN) を基にした autoencoder (AE) が用いられている．特に，AE の拡張である variational autoencoder (VAE) は与えられた画像に対する精密な確率モデルを構築する [Kingma 14]．DNN はデータに従い高次の特徴を学習することで，様々なタスクでとても良い結果を得ている．しかし，DNN は誤って認識しているサンプルに対しても高い尤度を割り当てることが分かっている [Goodfellow 15].

このような背景から，我々は VAE を用いたセグメンテーションに対する新たな異常度を提案する．VAE がサンプルの尤度を推定することから，いくつかの既存研究は元々のサンプルと推定された平均の二乗誤差を異常度として用いている [Chen 18]．代わりに，我々はピクセルごとの条件付き尤度の近似を，偶然的不確実性 (aleatoric uncertainty [Kendall 15, Kendall 17, Kiureghian 09]) と正規化誤差の 2 つの項に分解し，サンプル単位での異常検知に関する既存研究 [Matsubara 18] に従い正規化誤差の項のみを VAE の異常度として用いることを提案する．

連絡先: 佐藤一輝，神戸大学 大学院システム情報学研究科，ksato@ai.cs.kobe-u.ac.jp

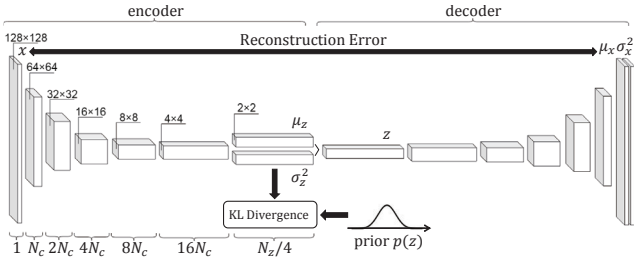

図 1: A diagram of the variational autoencoder (VAE) implemented on convolutional neural networks (CNNs).

我々は，頭部 MRI のデータセットを用いた実験により，提案した異常度の評価を行う．この実験では，脳腫瘍および脳卒中後の病変を教師なし学習で検出する．実験結果により提案手法が比較手法よりも良い性能を持つことを示し，また定性的比較により期待した通り偽陽性が減少することを示す．

2. 関連研究

autoencoder (AE) は，一般に 2 つの構成要素 encoder と decoder から成る DNN の一種である．encoder は入力したサンプル x を低次元の中間変数 z に写像する．decoder は中間変数 z を受け取り，元のサンプル x を再生するように再構成 $\tilde{x}$ を出力する．通常，AE で最小化される目的関数は，サンプル x と再構成 $\tilde{x}$ の mean-squared-error (MSE) である．異常度としては，サンプル単位での異常検知では MSE を，ピクセル単位での異常検知ではピクセルごとの二乗誤差を用いることが出来る．AE は潜在変数 z を推論して x を単位行列の分散共分散行列をもつ正規分布の事後分布として出力する生成モデルとして見なせる．

variational autoencoder (VAE，図 1 参照) は更に明示的に事後分布のモデルを構築する [Kingma 14]．データ空間 $\mathcal{X} \subset \mathbb{R}^{N_x}$ 内のサンプル x について，潜在空間 $\mathcal{Z} \subset \mathbb{R}^{N_z} (N_z < N_x)$ 内の潜在変数 z を持つ次のような確率モデルを考える．

$$p_\theta(x) = \int_z p_\theta(x|z)p(z)$$

変分推論を用いて，モデルエビデンス $\log p_\theta(x)$ は

$$\begin{aligned}\log p_\theta(x) &= \mathbb{E}_{q_\phi(z|x)}\left[\log \frac{p_\theta(x,z)}{p_\theta(z|x)}\right] \\ &= \mathbb{E}_{q_\phi(z|x)}\left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\right] + D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \\ &\geq \mathbb{E}_{q_\phi(z|x)}\left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\right] \\ &= -D_{KL}(q_\phi(z|x)||p(z)) + \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] \\ &=: -\mathcal{L}(x)\end{aligned} \tag{1}$$

という下界を取る．ここで，$-\mathcal{L}(x)$ は変分下界 (evidence lower bound; ELBO) と呼ばれる．VAE では ELBO を目的関数として最大化する．ELBO は対数尤度そのものではないが，対数尤度に収束すると考えることが出来る．

DNN の柔軟性と誤差逆伝播による最適化を利用するため，VAE は確率モデル $q_\phi(z|x)$ と $p_\theta(x|z)$ を reparameterization trick を用いてそれぞれ encoder と decoder として実装する．すなわち，encoder と decoder はそれぞれ対応する確率モデルの点推定などを出力するのではなく，それらのパラメータを出力する．DNN では式 (1) の期待値を求めることが困難であるため，モンテカルロサンプリングにより近似する．原論文 [Kingma 14] の実装では，潜在変数 z は訓練中に一度ずつサンプリングされる．

3. 提案手法

本研究では，サンプルの各要素 (ピクセル) は連続的な変数であり，対角な分散共分散行列を持つ多変量正規分布としてモデル化する．$f(\cdot|\mu,\Sigma)$ を平均 μ，分散共分散行列 Σ の多変量正規分布とする．このとき負の ELBO $\mathcal{L}(x)$ は

$$\begin{aligned}\mathcal{L}(x) &= \int_z f(z;\mu_z,\mathrm{diag}(\sigma_z^2))\log\frac{f(z;\mu_z,\mathrm{diag}(\sigma_z^2))}{f(z;\mathbf{0},I)} \\ &\quad - \mathbb{E}_{q_\phi(z|x)}\left[-\log f(x;\mu_x,\mathrm{diag}(\sigma_x^2))\right] \\ &= \sum_j \frac{1}{2}(-\log\sigma_{z_j}^2(x) - 1 + \sigma_{z_j}^2(x) + \mu_{z_j}^2(x)) \\ &\quad + \mathbb{E}_{q_\phi(z|x)}\left[\textstyle\sum_i \frac{1}{2}\log 2\pi\sigma_{x_i}^2(z) + \sum_i \frac{(\mu_{x_i}(z)-x_i)^2}{2\sigma_{x_i}^2(z)}\right]\end{aligned} \tag{2}$$

と書き直すことが出来る．

画像における異常部分 (ピクセル) の検出を目的とするので，ここでは潜在変数 z が与えられたもとでの画像 x の事後分布に着目する．いま，x を対角な分散共分散行列をもつ多変量正規分布としてモデル化しているため，個々のピクセルについてこれを扱うことができる．ピクセル x_i の負の対数尤度 $\mathcal{L}(x_i;x)$ の近似は

$$\begin{aligned}\mathcal{L}(x_i;x) &= \log f(x_i;\mu_{x_i}(x),\sigma_{x_i}^2(x)) \\ &= \underbrace{\frac{1}{2}\log 2\pi\sigma_{x_i}^2(x)}_{\substack{\text{aleatoric uncertainty}\\ \mathcal{U}(x_i;x)}} + \underbrace{\frac{(\mu_{x_i}(x)-x_i)^2}{2\sigma_{x_i}^2(x)}}_{\substack{\text{normalized error}\\ \mathcal{E}(x_i;x)}}\end{aligned} \tag{3}$$

と変形できる．このピクセルごとの負の対数尤度 $\mathcal{L}(x_i;x)$ は異常度として用いることが出来る．

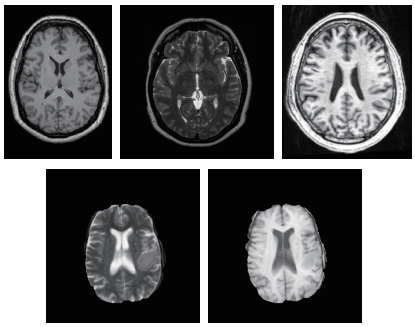

図 2: Examples images in datasets. (upper row) IXI-T1w, IXI-T2w, and ATLAS-T1w. (bottom row) BraTS-T2w, and BraTS-T1w.

第一項 $\mathcal{U}(x_i;x)$ は事後分布の対数分散 (log-variance) に関するものであり，第二項 $\mathcal{E}(x_i;x)$ は分散により正規化された二乗誤差である．ピクセル x_i が二つの領域の境界上にあったり，ノイズが乗っていたり，複雑な形状をもつ領域に存在していたりする場合に，x_i は予測が困難になり，二乗誤差 $(\mu_{x_i}(x)-x_i)^2$ は大きくなる．そこで，VAE は分散 $\sigma_{x_i}^2$ を大きい値に仮定することで，正規化誤差 $\mathcal{E}(x_i;x)$ を抑えながら log-variance $\mathcal{U}(x_i;x)$ を増加させる．すなわち，VAE は log-variance $\mathcal{U}(x_i;x)$ と正規化誤差 $\mathcal{E}(x_i;x)$ を平衡させるものであり，$\mathcal{U}(x_i;x)$ はデータが本来持つばらつきに起因する不確かさを表す偶然的不確実性 (aleatoric uncertainty) としてみなすことが出来る．我々は対数尤度 $\mathcal{L}(x_i;x)$ から不確かさ $\mathcal{U}(x_i;x)$ を取り除いた $\mathcal{E}(x_i;x)$ を新たな異常度として提案する．

4. 評価実験

4.1 データセット

頭部磁気共鳴法 (MRI) とは，強力な磁場とパルス状の電波により特定の原子核の磁気共鳴を誘発し，その状態を調べることで脳組織を画像化する技術である．電波の放射間隔とエコー時間を変えることで，特定の部位を強調した複数の種類の画像 (T1, T2, Flair, etc.) を得ることが出来る．我々は提案手法を頭部 MRI データセットを用いて評価した．画像の例を図 2 に示す．

- **IXI dataset**: モデルの学習には IXI データセットを用いた．このデータセットには 579 人の健康な人から得られた T1 及び T2 強調の MRI が含まれている．T1 強調画像は 256 枚の 150×256 の大きさの水平断画像から成り，T2 強調画像は 130 枚または 136 枚の 256×256 の大きさの水平断画像から成る．
- **ATLAS-T1w**[Liew 18]: モデルの評価には，220 人の脳卒中発症者から得られた T1 強調画像を含む Anatomical Tracings of Lesions After Stroke (ATLAS) データセットを用いた．画像は 189 枚の 197×233 の大きさの水平断画像から成る．
- **BraTS**[Menze 15, Bakas 17]: 加えて，更なる評価のため 285 人の脳腫瘍患者から得られた T2 及び T1 強調の MRI を含む Brain Tumor Segmentation Challenge (BraTS) データセットを用いた．T2,T1 強調画像ともに 155 枚の 240×240 の大きさの水平断画像から成る．

データの前処理として，まず Brain Extraction Tools (BET) [Smith 02] を用いて脳以外の部位を除去した．ただ

表 1: Resultant ROC-AUCs.

Model	Score	ATLAS-T1w Hyper-Parameters	ATLAS-T1w AUC	BraTS-T2w Hyper-Parameters	BraTS-T2w AUC	BraTS-T1w Hyper-Parameters	BraTS-T1w AUC
Baseline	pixel intensity	—	0.612	—	0.811	—	0.549
modified GMM	$\lambda/(f(x_i)+\lambda)$	$\lambda = 0.001, N_z = 4$	<u>0.635</u>	$\lambda = 0.001, N_z = 4$	**0.798**	$\lambda = 0.001, N_z = 4$	0.550
AE	squared Error	$N_c = 16, N_z = 8$	0.588	$N_c = 16, N_z = 256$	0.566	$N_c = 16, N_z = 16$	0.545
DAE	squared Error	$N_c = 16, N_z = 8$	0.583	$N_c = 16, N_z = 256$	0.607	$N_c = 16, N_z = 16$	0.532
VAE	squared Error	$N_c = 32, N_z = 8$	0.611	$N_c = 16, N_z = 256$	0.733	$N_c = 32, N_z = 16$	0.554
	$\mathcal{L}(x_i; x)$	$N_c = 32, N_z = 8$	0.626	$N_c = 16, N_z = 256$	0.737	$N_c = 32, N_z = 16$	<u>0.561</u>
	$\mathcal{U}(x_i; x)$	$N_c = 32, N_z = 8$	0.404	$N_c = 16, N_z = 128$	0.520	$N_c = 16, N_z = 16$	0.436
VAE	$\mathcal{E}(x_i; x)$	$N_c = 32, N_z = 8$	**0.672**	$N_c = 16, N_z = 256$	<u>0.788</u>	$N_c = 32, N_z = 16$	**0.582**

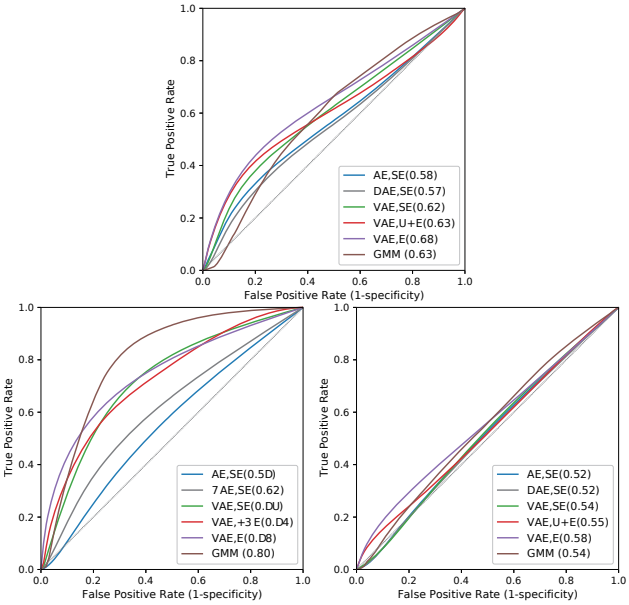

図 3: Receiver operating characteristic (ROC) curves of the models with the best hyper-parameters. The datasets are ATLAS-T1w (top), BraTS-T2w (bottom left), and BraTS-T1w (bottom right).

し，図 2 に示した通り，BraTS は除去済みのデータが配布されているためこの処理は不要であった．その後，脳の部分を切り抜き，同じサイズ 128×128 に揃えた．さらに，各画像のピクセルを脳の存在する領域の強度の中央値で割る正規化処理を施した．

4.2 モデル

AE 系のモデルとしては，VAE に加え比較対象として AE と denoising autoencoder(DAE) を実装した．DAE は入力画像 x に平均 0, 標準偏差 0.5 のガウス雑音を加えて学習する AE である [Chen 18]．これらの encoder は 4×4 のカーネルと 2 のストライドを持つ 6 層の畳み込み層から成る．n 層目 $(n < 6)$ の後の特徴マップは $N_c \times 2^{n-1}$ チャネルを持ち，バッチ正規化と ReLU 活性化関数が続く．最後の畳み込み層は 3×3 のカーネルと 1 のストライドを持ち，$2 \times 2 \times \frac{N_z}{4}$ のサイズの特徴マップに達する．最後の特徴マップは長さ N_z の特徴ベクトルに変形される．これは潜在変数 z である．VAE では，encoder が出力する長さ N_z の特徴ベクトルのうち半数は平均ベクトル μ_z として，残る半数は対数分散ベクトル $\log \sigma_z^2$ として，それぞれ用いられる．これらは，多変量正規分布である事後分布 $q_\phi(z|x)$ のパラメータを表現する．decoder は encoder と対を成す構造を持つ．VAE では出力層が 2 チャネルの画像に，これらが多変量正規分布である事後分布 $p_\theta(x|z)$ の平均 μ_x と対数分散 $\log \sigma_x^2$ を表す．

AE 系のモデルの最適化アルゴリズムとしては Adam ($\alpha = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) を 10^{-4} の重み減衰で用いた．潜在空間の次元数を $N_z \in \{8, 16, 32, 64, 128, 256, 512\}$ で，チャネル数を $N_c \in \{8, 16, 32\}$ で探索した．AE と DAE では，最小化を行う目的関数には MSE を，異常度にはピクセルごとの二乗誤差を用いた．VAE には 2. 節で述べた元々の目的関数を用いた．VAE の異常度として，比較のためピクセル単位の二乗誤差および対数尤度の近似 $\mathcal{L}(x_i; x)$，そして提案する異常度 $\mathcal{E}(x_i; x)$ を用いてそれぞれ評価を行った．

更なる性能評価のため，代表的な生成モデルの一つである混合ガウスモデル (GMM) を基にしたモデル [Van Leemput 01] を実装した．元の GMM では各ピクセルの分布を N_z 個の正規分布 $f(x_i; \theta_k), k \in \{1, \ldots, N_z\}$ の混合としてモデル化するのに対して，このモデル (以降では modified GMM と呼ぶ) では正規分布と一様分布の混合でモデル化する．これにより，学習データに含まれる外れ値に対するモデルの頑健性が増す．あるピクセルが一様分布から生成されたものである事後確率 $\sum_{k=1}^{N_z} \frac{\lambda}{f(x_i;\theta_k)+\lambda}$ を異常度として用いる．一様分布の含まれる度合いを示すハイパーパラメータ λ を $\lambda \in \{0.1, 0.01, 0.001, 0.0001\}$ で探索した．また，隠れクラスの数 N_z は，知られている脳組織の分類 {WM,GM,CSF} に背景のクラスを加えた 4 とした．

5. 結果と考察

性能評価のため，我々は receiver operating characteristic (ROC) 曲線を描画した．ROC 曲線とは，縦軸を陽性率，横軸を偽陽性率として，異常判定の閾値を変えながらプロットしたものである．さらに，ROC 曲線の下部の面積 (ROC-AUC) を求めた．最も良い ROC-AUC を得たハイパーパラメータとその ROC-AUC を表 1 に，対応する ROC 曲線を図 3 に示す．いずれのデータセットについても，提案した異常度 $\mathcal{E}(x_i; x)$ を用いた VAE は $\mathcal{L}(x_i; x)$ を用いた場合よりも良い性能を示した．特に ATLAS-T1w においては，他のどの比較手法よりも性能が上回った．しかし，BraTS-T2w に対しては，modified GMM が他の手法よりも良い性能を示し，さらにピクセルの強度をそのまま異常度として用いた場合に最も良い ROC-AUC が得られた．一般に，T2 強調の MRI は水分が多く含まれる部位が明るいピクセルとして映るため，診断するモデルはある閾値を超えたピクセルを腫瘍として容易に検知できる．それゆえ，GMM が良い性能を示したのは，モデルが単純であることに起因すると考えられる．一方で，BraTS-T1w に対しては提

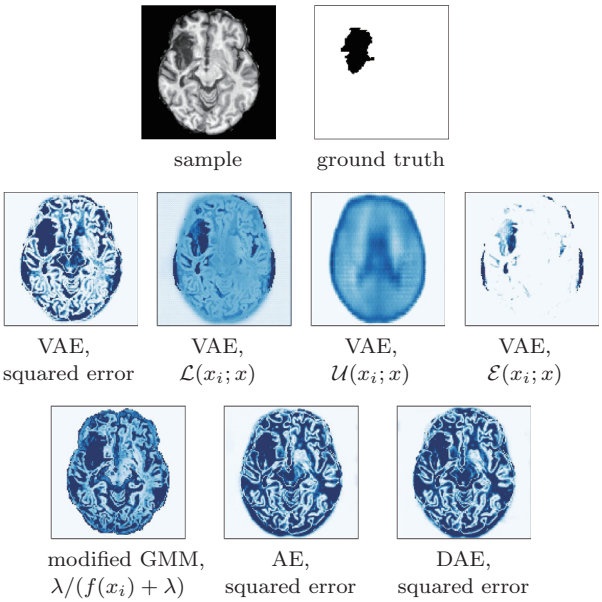

図 4: A sample from the ATLAS-T1w dataset, the ground truth of anomalous regions, and the corresponding heat maps of the anomaly scores.

案した $\mathcal{E}(x_i;x)$ で最も良い結果を得た．T1 強調画像では T2 強調画像に比べ脳の構造が強調されるため，診断にはより意味論的かつ構造的な分析を施す必要がある．従って，提案した異常度 $\mathcal{E}(x_i;x)$ は他の AE 系の手法や GMM に比べて意味論的な異常のセグメンテーションにおいて特に良い性能を持つ．

図 4 に，ATLAS-T1w のサンプル，異常部分の正解アノテーション，各モデルで得られた異常度を可視化したものを示す．VAE の二乗誤差は脳の複雑な構造に対応できておらず，これらの部位に対して高い異常度を示している．AE および DAE の二乗誤差，また GMM についても同様である．これらの異常度は意味論的な異常よりもむしろそのピクセルの強度の絶対値の大きさに敏感である．VAE で対数尤度の近似 $\mathcal{L}(x_i;x)$ を用いた場合は，異常部分に高い異常度を示したが，それ以外の部位においてもある量の異常度を示した．これは VAE が，複雑で個人差のある部位の再構成誤差を緩和するために $\mathcal{U}(x_i;x)$ として高い値を出力しているためである．$\mathcal{U}(x_i;x)$ の作用により，提案した正規化誤差の項 $\mathcal{E}(x_i;x)$ は異常部分を選択的に検出することが出来る．以上から，正規化誤差 $\mathcal{E}(x_i;x)$ は与えられた画像の構造における複雑性に対して頑健であり，意味論的な異常を良く検出できることが結論づけられる．

6. 結論

本研究では，教師なし異常部分検知における VAE の新しい異常度を提案した．この異常度は VAE の目的関数から log-variance 項を取り除くことで得られた．log-variance 項が異常よりもデータの複雑性に反応することから，残った項は複雑性に対して頑健になる．我々は提案した異常度を頭部 MRI における病変箇所のセグメンテーションにより評価した．結果として，複雑性に対する頑健さのため，提案した異常度はこれらのデータセットに対して良い性能を示した．

本研究は総務省 SCOPE(受付番号 172107101) の委託を受けて行われた．

参考文献

[Bakas 17] Bakas, S., Akbari, H., Sotiras, A., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, *Scientific Data*, Vol. 4, No. March, pp. 1–13 (2017)

[Chen 18] Chen, X., Pawlowski, N., Rajchl, M., et al.: Deep Generative Models in the Real-World: An Open Challenge from Medical Imaging, *arXiv*, pp. 1–10 (2018)

[Goodfellow 15] Goodfellow, I. J., Shlens, J., and Szegedy, C.: Explaining and Harnessing Adversarial Examples, in *International Conference on Learning Representations (ICLR)*, pp. 1–11 (2015)

[Havaei 17] Havaei, M., Davy, A., Warde-Farley, D., et al.: Brain tumor segmentation with Deep Neural Networks, *Medical Image Analysis*, Vol. 35, pp. 18–31 (2017)

[Kendall 15] Kendall, A., Vijay Badrinarayanan, , Cipolla, R., et al.: Bayesian SegNet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding, in *arXiv* (2015)

[Kendall 17] Kendall, A. and Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, *Advances in Neural Information Processing Systems (NIPS)* (2017)

[Kingma 14] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, in *International Conference on Learning Representations (ICLR)*, pp. 1–14 (2014)

[Kiureghian 09] Kiureghian, A. D. and Ditlevsen, O.: Aleatory or epistemic? Does it matter?, *Structural Safety*, Vol. 31, No. 2, pp. 105–112 (2009)

[Liew 18] Liew, S.-L., Anglin, J. M., Banks, N. W., et al.: A large, open source dataset of stroke anatomical brain images and manual lesion segmentations, *Scientific Data*, Vol. 5, p. 180011 (2018)

[Matsubara 18] Matsubara, T., Tachibana, R., and Uehara, K.: Anomaly Machine Component Detection by Deep Generative Model with Unregularized Score, in *International Joint Conference on Neural Networks (IJCNN)* (2018)

[Menze 15] Menze, B. H., Jakab, A., Bauer, S., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), *IEEE Transactions on Medical Imaging*, Vol. 34, No. 10, pp. 1993–2024 (2015)

[Smith 02] Smith, S. M.: Fast robust automated brain extraction, *Human brain mapping*, Vol. 17, No. 3, pp. 143–155 (2002)

[Van Leemput 01] Van Leemput, K., Maes, F., Vandermeulen, D., et al.: Automated segmentation of multiple sclerosis lesions by model outlier detection, *IEEE Transactions on Medical Imaging*, Vol. 20, No. 8, pp. 677–688 (2001)