Proces rekonstrukcji obrazu tomograficznego w oparciu o sieć Variational Autoencoder

The process of reconstruction in a CT image based on an variational autoencoder network

Jolanta Podolszańska

Politechnika Częstochowska, Wydział Inżynierii Mechanicznej i Informatyki, Katedra Inteligentnych Systemów Informatycznych, Al. Armii Krajowej 21
42-201 Częstochowa, tel. +48 34 325 05 46, e-mail: jolanta.podolszanska@pcz.pl

Wprowadzenie

Pierwsze modele sieci neuronowych powstały już w 1943 roku, ale dopiero teraz możliwy stał się ich rozwój. Wczesne modele sieci neuronowych zapoczątkowały pierwszą falę rewolucji z nimi związanych. Chociaż na początku pokładano wielkie nadzieje i objawiało się to obietnicami i przewidywaniami, a nie konkretnymi projektami z ich zastosowaniem, to z biegiem czasu zaczęło się to wszystko zmieniać. Wraz z rozwojem technologii sieci neuronowe znów wróciły do łask. Jednak nie można lekceważyć faktu, że sam pomysł robienia czegokolwiek, co jest oparte na sieciach neuronowych, początkowo spotkał się z rozczarowaniem. Świat naukowy przez ponad 10 lat ignorował postępy np. z algorytmem wczesnej propagacji.

Proces rekonstrukcji obrazu tomograficznego

Za pojęciem rekonstrukcji obrazu kryje się pewna metoda, która ma na celu odtworzenie oryginalnego obrazu. Sama rekonstrukcja w ujęciu tomografii komputerowej jest procesem matematycznym, który generuje obrazy na podstawie danych projekcji rentgenowskiej uzyskanej pod wieloma kątami wokół ciała pacjenta. Dziś wręcz pożądana jest rekonstrukcja obrazów z jak najmniejszym szumem bez utraty rozdzielczości przestrzennej, także z wykorzystaniem mniejszej dawki promieniowania.

Istnieją dwie główne kategorie metod rekonstrukcji: iteracyjna i analityczna. Rekonstrukcja iteracyjna odtwarza obrazy przy pomocy iteracyjnej funkcji celu. Rekonstrukcja iteracyjna wpływa także na redukcję artefaktów obrazów (utwardzenie wiązki, metalowe artefakty). Metoda analityczna najczęściej ma postać filtrowanej projekcji wstecznej (FBP). Metoda ta wykorzystuje filtr na dane projekcji przed odwzorowaniem danych na przestrzeń obrazu. Algorytmy rekonstrukcyjne są testowane na matematycznym modelu głowy Shepp-Logana (Ryc. 1) w celu ich przetestowania.

Rozwój algorytmów rekonstrukcyjnych stał się łatwiejszy dzięki zastosowaniu standardowych fantomów o znanych właściwościach. Model Shepp-Logan zawiera elipsy o różnych właściwościach absorpcyjnych, przypominając swoim kształtem

62

Streszczenie

Artykuł ma na celu zapoznanie się z rekonstrukcją i odszumianiem obrazu za pomocą sieci neuronowej typu VAE (Variational Auto-Encoder). W pracy zostanie dokonana analiza porównawcza pod kątem błędów rekonstrukcji i występujących na obrazie anomalii. Posłużono się zbiorem obrazów TK mózgu (Visible Female CT), aby pokazać, jak wygląda rekonstrukcja i odszumianie metodą Variational Autoencoder.

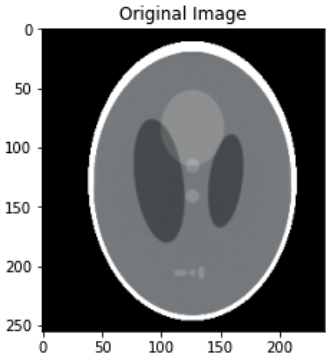
Słowa kluczowe: tomografia komputerowa, obrazowanie medyczne, autokoder wariacyjny, przetwarzanie obrazu

Abstract

This paper aims to learn about image reconstruction and de-noising using Variational Encoder (VAE) neural network. The paper will make a comparative analysis in terms of reconstruction errors and anomalies present in the image. A collection of brain CT images (Visible Female CT) is used to show how reconstruction and de-noising by Variational Autoencoder method.

Key words: computed tomography, medical imaging, variational autoencoder, vae, image processing

otrzymano / received:
26.01.2021
poprawiono / corrected:
01.02.2021
zaakceptowano / accepted:
12.02.2021



Ryc. 1 *Model Shepp-Logan*
Źródło: Implementacja własna w języku Python.

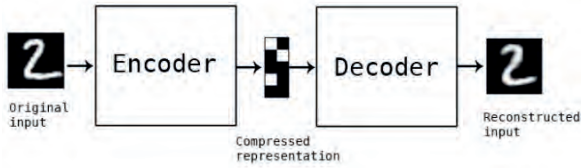
zarys głowy. Umożliwia to generowanie obrazów o różnych rozmiarach oraz zmianę generowanych projekcji. Istnieje także wariant rozszerzonego modelu Sheppa-Logana zawierającego 12 elips.

Metoda Variational Autoencoder

Idea Variational Autoencodera jest najmniej podobna do innych znanych autoencoderów, ale została silnie zakorzeniona w metodach wariacyjnego i graficznego modelu. W innych autoencoderach dane wejściowe były mapowane na ustalony wektor, tutaj jest zupełnie inaczej. Zamiast zwykłego mapowania danych wejściowych na ustalony wektor, są one mapowane bezpośrednio do rozkładu. Dany rozkład ρ_θ został sparametryzowany przez θ. Zależność pomiędzy wejściem danych a utajonym wektorem kodowania można zdefiniować za pomocą kilku zależności. Znając rzeczywisty parametr θ^* rozkładu, możemy wygenerować próbkę, która wygląda jak rzeczywisty punkt danych $x^{(i)}$. W tym celu pobieramy próbkę z $z^{(i)}$ z wcześniejszego rozkładu $\rho_{\theta^*}z$, następnie wartość $x^{(i)}$ jest generowana z rozkładu warunkowego $\rho_{\theta^*}z\,(x|z=z^{(i)})$.

Sieć dekodera definiuje rozkład warunkowy obserwacji $\rho(x|z)$, która pobiera utajoną próbkę z jako dane wejściowe i wyprowadza parametry warunkowego rozkładu obserwacji. Została zamodelowana utajona przestrzeń przez $\rho(z)$ jako jednostkę Gaussa. Zostały użyte dwie warstwy konwolucyjne, a dzięki ich połączeniu powstaje w pełni połączona warstwa. Po połączeniu warstwy powstają kolejno trzy warstwy transpozycji splotu. W tym podejściu odradza się używania stosowania normalizacji, gdyż dodatkowa stochastyczność może pogorszyć niestabilność modelu, nie wliczając w to stochastyczności z próbkowania.

Stosowanie metody Autoencoder można spotkać w tomografii komputerowej z niską dawką promieniowania. Nie jest żadną tajemnicą, że statystyczne modelowanie charakterystyki jest trudnym zadaniem. Obrazy w niektórych przypadkach nie są w stanie wyeliminować zakłóceń obrazu, stosując niską dawkę promieniowania, a jednocześnie zachowując wszystkie szczegóły rekonstrukcyjne. Zaproponowane podejście pokazuje interesujące wyniki rekonstrukcji obrazu z wykorzystaniem techniki autoencodera typu VAE.



Ryc. 2 *Schemat obrazu wejściowego/wyjściowego z przykładowego zestawu danych MNIST do autodekodera*
Źródło: https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726

Dane treningowe

Jako dane treningowe użyto obrazy tomograficzne głowy kobiety o dobrze znanym już rozmiarze 256 x 256 pikseli. Encoder pobiera partie obrazów, a następnie wytwarza dwa ukryte wektory: μ i δ. Sieć została zaprojektowana tak, aby co druga warstwa zmniejszała swoją rozdzielczość o połowę. Otrzymujemy w pełni połączoną warstwę z 1024 kanałami i jądrem o rozmiarze 8 x 8. W taki sposób zostały połączone dwie warstwy z 512 kanałami. Warstwy przewidują jeden z ukrytych wektorów μ i σ. Za pomocą procesu reparametryzacji został dodany szum do rozkładu utajonego. Warto podkreślić, że tylko część dekodera bierze czynny udział w generowaniu pełnowymiarowego obrazu wyjściowego. Wykorzystano moduł Conv2D, który stosuje splot 2D do sygnału wejściowego złożonego z kilku płaszczyzn wejściowych. Zatem wartość wyjściową z rozmiarem wejściowym można opisać jako:

$$out\left(N_i, C_{outj}\right) = bias\left(C_{outj}\right) + \sum_{k=0}^{C_j - 1} weight\left(C_{out}, k\right) * input\left(N_i, k\right)$$

* – poprawny operator korelacji krzyżowej,
N– wielkość partii,
C – liczba kanałów,
H – wysokość płaszczyzn wejściowych wyrażona w pikselach,
W – szerokość wyrażona w pikselach.

Moduł Convd2D zarządza połączeniami pomiędzy wejściami i wyjściami, dane powinny być pogrupowane. W pierwszej grupie dane z wejścia są połączone ze wszystkimi wyjściami. W drugiej grupie sieć posiada tzw. warstwy konwolucyjne, które znajdują się obok siebie. Każda z nich widzi połowę kanałów wejściowych i wpływa na połowę kanałów wyjściowych. Na końcu są dopiero łączone. W pozostałych grupach każdy kanał wejściowy jest zwijany z własnym zestawem filtrów o następującej wielkości:

$$\left[\frac{out\_channels}{in\_channels}\right]$$

gdzie:
out_channels – liczba kanałów wytworzona przez konwulsję,
in_channels – liczba kanałów na obrazie wejściowym,
groups – liczba zablokowanych połączeń z kanałów wejściowych do wyjściowych.

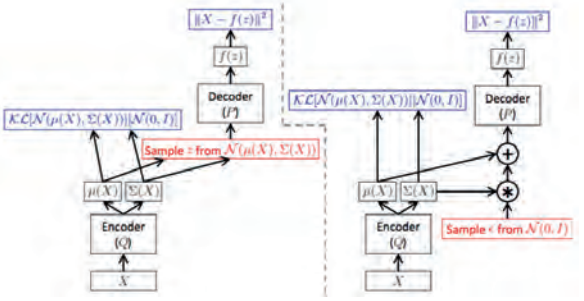
O procesie zwijania wgłębnego (*depthwise convolution*) mówimy, kiedy

*groups == out_channels && in_channels == K * in_channels*

przy założeniu, że K jest dodatnią liczbą całkowitą, dla danych wejściowych o wielkości (N, C_{in}, H_{in}, W_{in}) zwijanie wgłębne

z multiplikatorem głębinowym *K* mogą być konstruowane za pomocą argumentów *in_channels* = C_{in}, *out_channels* = C_{in} x *K*,..., *groups* = C_{in}.

W pełni połączona warstwa Conv2D mapuje interpretowany 512-wymiarowy wektor utajony do stanu ukrytego, który posiada 1024 kanały. Następnie dokonuje się transpozycja zwoju z jądrem, a następnie Conv2D z jądrem na przemian z blokami ConvTransponse2D. Ta operacja ma na celu zwiększenie rozdzielczości mapowanych cech. Oprócz wygenerowania obrazu wyjściowego sieć zwraca przewidywanie dla ukrytych wektorów μ i σ, które są wykorzystywane do oszacowania straty.



Ryc. 3 *Architektura sieci VAE*
Źródło: https://blog.qure.ai/notes/using-variational-autoencoders.

Funkcja straty trenowana jest poprzez maksymalizację dolnej granicy dowodów (ELBO) na marginalnym logarytmicznym prawdopodobieństwie:

$$\log p(x) \geq ELBO = E_{q(z|x)}\left[\log \frac{p(x,z)}{q(z|x)}\right]$$

Badanie

Badanie ma na celu przedstawienie rekonstrukcji obrazu za pomocą wcześniej wspomnianej metody. Zastosowano tutaj sieć typu enkoder-dekoder (z koderem i dekoderem) z naciskiem na wykorzystanie metody VAE. W warstwie ukrytej nie bez powodu znajdują się bardziej złożone filtry. Otóż warstwa ukryta jest bardziej zagęszczona, co oznacza, że trudno przewidzieć rozkład wartości w tej przestrzeni. Niemniej jednak jest to dobra właściwość dla systemów kompresji. Minusem tej metody może być sam fakt, że jeśli pomiędzy klastrami występują luki, dekoderowi trudno będzie wygenerować coś użytecznego, ze względu na ubytki w wiedzy o tym zjawisku. W tym badaniu zostanie zastosowane podejście *Variational Autoencoder*, które uczyni warstwę ukrytą bardziej przewidywalną. Przestrzeń będzie bardziej ciągła, przez co zmusi utajone zmienne do stania się bardziej rozproszonymi. Dzięki temu rozwiązaniu sieć uzyska kontrolę nad utajoną przestrzenią. Pozostaje sam problem bezpośredniego przekazywania wartości do dekodera. *VAE* ten problem rozwiązał zupełnie od innej strony. Zanim wartości zostaną przekazane do dekodera są używane do obliczenia średniej i odchylenia standardowego. Wejście dekodera jest wstępnie próbkowane z odpowiedniego rozkładu normalnego. Podczas treningu następuje pewne wymuszenie, aby rozkład normalny był zbliżony w odniesieniu do standardowego rozkładu normalnego. Autoencodery wariancyjne pozwalają modelować dane wejściowe zarówno w oparciu o ukrytą zmienną *z*, jak i dodatkowe metadane.

Wstępne wyniki przeanalizowano na podstawie zestawu testowego MNIST (Ryc. 4), a następnie zastosowano zmodyfikowany algorytm na zestawie testowym obrazów tomograficznych głowy. Obrazy zostały pobrane losowo z wyuczonej sieci z warstwy ukrytej. VAE został przeszkolony bez nadzoru w zakresie rekonstrukcji i straty. Zarysy poszczególnych organów są wyraźnie widoczne. Można jednak zauważyć, że pomimo realistycznie wyglądających obrazów pojawia się tutaj jeden problem, mianowicie są one zamazane i brakuje im drobnych szczegółów.

Sam proces uczenia nienadzorowanego polega na optymalizacji rekonstrukcji na wejściu i w generowanych obrazach. Zostało tutaj poruszone pojęcie dywergencji Kullback-Leiblera, które określa rozbieżności pomiędzy dwoma rozkładami prawdopodobieństwa. Dywergencja jest określona osobno dla rozkładu dyskretnego i dla rozkładu ciągłego.

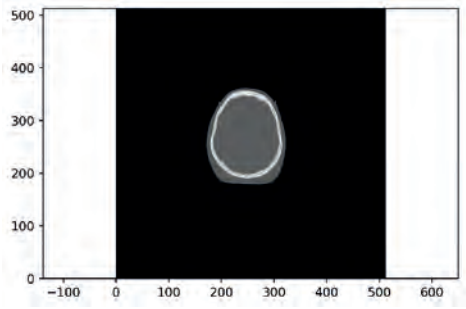
Ryc. 4 *Rekonstrukcja obrazu za pomocą techniki VAE na przykładzie datasetu MNIST*
Źródło: Wykonanie własne z użyciem datasetu MNIST http://yann.lecun.com/exdb/mnist/.

Rozkład może reprezentować zmianę lub różnicę w ilości informacji pomiędzy dwoma rozkładami. Takie podejście zastosowano także w przypadku obrazów syntetycznych. Zostały pobrane losowe obrazy głowy (Ryc. 5) i zastosowano tutaj podejście VAE (Ryc. 6). Można zauważyć, że pojawia się tutaj problem rozmazania obrazu, gdzie w tomografii komputerowej nie ma na to miejsca.

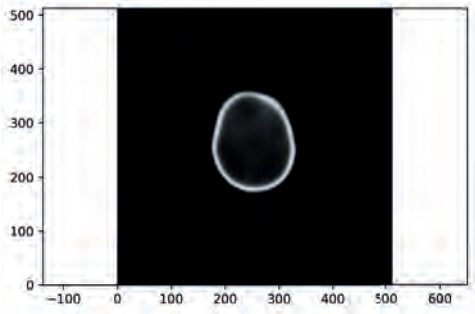
Zastosowanie metody VAE przy rekonstrukcji obrazów syntetycznych może sprawiać pewne problemy, jeśli skupimy się tutaj

```
Epoch: 1, Test set ELBO: -181.21673583984375, time elapse for current epoch: 15.664434671401978
Epoch: 2, Test set ELBO: -172.9552459716797, time elapse for current epoch: 14.935616731643677
Epoch: 3, Test set ELBO: -169.78457641601562, time elapse for current epoch: 14.81468653678894
Epoch: 4, Test set ELBO: -167.18495178222656, time elapse for current epoch: 15.011104106903076
Epoch: 5, Test set ELBO: -165.34561157226562, time elapse for current epoch: 15.33605670928955
Epoch: 6, Test set ELBO: -163.9119873046875, time elapse for current epoch: 15.409557580947876
Epoch: 7, Test set ELBO: -162.30105590820312, time elapse for current epoch: 15.428087711334229
Epoch: 8, Test set ELBO: -161.26866149902344, time elapse for current epoch: 15.50120735168457
Epoch: 9, Test set ELBO: -160.6299591064453, time elapse for current epoch: 15.590165615081787
Epoch: 10, Test set ELBO: -159.87612915039062, time elapse for current epoch: 15.611324787139893
```

Ryc. 5 *Trening sieci posiadającej parametry train_size = 60 000, batch_size = 32, test_size = 10 000, epochs = 10*
Źródło: Wykonanie własne środowisko Spyder.



Ryc. 6 *Oryginalny obraz tomograficzny ze zbioru Visible Female CT datasets head section*
Źródło: https://mri.radiology.uiowa.edu/visible_human_datasets.html.



Ryc. 7 *Rekonstrukcja obrazu z zastosowaniem algorytmu VAE*
Źródło: Wykonanie w środowisku Spyder.

na typowych algorytmach rekonstrukcyjnych. Uzyskane obrazy często są rozmazane i potrzebne jest opracowanie metody, która będzie w stanie „wyostrzyć" elementy rozmazane.

Napotkane problemy

Klasyczne autoencodery nie nadają się do generowania obrazów syntetycznych, stosując klasyczne algorytmy rekonstrukcyjne. Można to już zauważyć na etapie treningu kodera i sieci dekoderów. Próbkowanie przestrzeni ukrytej uniemożliwia otrzymanie realistycznych obrazów. Problem został rozwiązany poprzez nauczenie sieci funkcji gęstości prawdopodobieństwa i ograniczenie do wielowymiarowego rozkładu normalnego.

Rekonstrukcja za pomocą VAE pozostawia wiele do życzenia, zwłaszcza przy próbie rekonstrukcji mniejszych struktur, które są kluczowe w obrazie tomograficznym. Jednak trzeba przyznać, że sam proces pozbycia się szumów na obrazie i jego wyostrzenie może się polepszyć. O ile VAE jest często poruszanym pojęciem w artykułach naukowych, tak trzeba powiedzieć także o wadach tego rozwiązania. Pierwsza trudność leży w identyfikacji różnorodnych danych. Drugim problemem może być uzyskanie odpowiednio wyostrzonego obrazu. Odpowiednie zoptymalizowanie modelu może zagwarantować doskonałą rekonstrukcję oraz usunąć szum generowany przez zbyt dużą stałą wariancję. Korzystając z możliwości kompresji, dochodzi do zmniejszenia realizmu surowego dekodera. Dzięki temu rozwiązaniu jest nadzieja na uzyskanie bardziej precyzyjnej rekonstrukcji, uwzględniając nawet drobne struktury prześwietlanego narządu. Warto też wspomnieć, że słaba aproksymacja danych powoduje złożone problemy całej sieci typu VAE, zaciemniając wpływ nowych przestrzeni ukrytych.

Podsumowanie

W tym artykule został poruszony problem rekonstrukcji obrazu za pomocą VAE. Można dojść do różnych wniosków, zarówno tych dobrych, jak i złych. Na pewno istnieją lepsze metody rekonstrukcji niż te z użyciem VAE, ale można z nadzieją myśleć, że w przyszłości się to zmieni. Gdyby wszyscy badacze skupili się tylko na jednej metodzie, to z pewnością nie moglibyśmy mówić o jakimkolwiek przełomie w żadnej dziedzinie. Z pewnością nad algorytmami trzeba jeszcze popracować zanim wdroży się je w tomografach komputerowych. Istnieje bowiem nadzieja, że w tym polu z czasem będzie się coś zmieniać. Wskazuje na to wiele prac badawczych, które zostały opublikowane w ostatnim czasie. Miejmy jednak nadzieję, że idzie to ku lepszemu i już całkiem niedługo będzie można mówić o wielkim przełomie z użyciem sieci VAE.

Piśmiennictwo

1. R. Cierniak: *X-Ray Computed Tomography in Biomedical Engineering*, Springer, 2011, doi: 10.1007/978-0-85729-027-4.
2. M. Nishio, Ch. Nagashima, S. Hirabayashi, A. Ohnishi, K. Sasaki, T. Sagawa, M. Hamada, T. Yamashita: *Convolutional auto-encoder for image denoising of ultra-low-dose CT*, Heliyon, 3(8), 2017.
3. H. Chen et al.: *Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network*, IEEE Transactions on Medical Imaging, 36(12), 2017, 2524-2535, doi: 10.1109/TMI.2017.2715284 (2017).
4. T. Inglot: *Teoria informacji a statystyka matematyczna*, Mathematica Applicanda, 42 (1), 2014, 115-115, DOI: 10.14708/ma.v42i1.521, data dostępu: 14.10.2020.
5. S.U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, T. Çukur: *Image synthesis in multi-contrast mri with conditional generative adversarial networks*, IEEE Transactions on Medical Imaging, 38(10), 2019, 2375-2388.
6. V. Berger, M. Sebag: *Variational Auto-Encoder: not all failures are equal*, 2020.
7. N. Pawlowski, M.C.H. Lee, M. Rajchl, S. McDonagh, E. Ferrante, K. Kamnitsas, S. Cooke, S. Stevenson, A. Khetani, T. Newman, F. Zeiler, R. Digby, J.P. Coles, D. Rueckert, D.K. Menon, V.F.J. Newcombe, B. Glocker: *Unsupervised Lesion Detection in Brain CT using Bayesian Convolutional Autoencoders*, Medical Imaging with Deep Learning Conference, 2018.
8. J. An, S. Cho: *Variational autoencoder based anomaly detection using reconstruction probability*, [in:] SNU Data Mining Center, Tech. Rep., 2015.
9. Y. Gal, Z. Ghahramani: *Bayesian convolutional neural networks with Bernoulli approximate variational inference*, [in:] arXiv preprint arXiv:1506.02158, (2015).
10. Ch. Louizos, M. Welling: *Multiplicative Normalizing Flows for Variational Bayesian Neural Networks*, [in:] arXiv preprint arXiv:1703.01961, 2017.
11. Y. Bando, K. Sekiguchi, K. Yoshii: *Adaptive Neural Speech Enhancement with a Denoising Variational Autoencoder*, INTERSPEECH 2020 Conference, Shanghai, China, 2020.
12. D. Zimmerer, J. Petersen, S.A.A. Kohl, K.H. Maier-Hein: *A Case for the Score: Identifying Image Anomalies using Variational Autoencoder Gradients*, Image and Video Processing, 2019.
13. Zbiór danych Visible Female CT dataset head section, https://mri.radiology.uiowa.edu/visible_human_datasets.html