

Redes Generativas Adversarias 3D de Crecimiento Progresivo para generación de mapas de atenuación de PET

3D-Progressive Growing Generative Adversarial Networks for PET attenuation map generation

Presentación: 08/10/2020

Doctorando:

Ramiro Germán Rodríguez Colmeiro

Grupo de Inteligencia Artificial y Robótica (GIAR), Unidad de Investigación y Desarrollo de las Ingenierías (UIDI), Facultad Regional Buenos Aires, Universidad Tecnológica Nacional - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) - Argentina / GAMMA3 (UTT-INRIA), Université de Technologie de Troyes - Francia
rodriguez.colmeiro@gmail.com

Director/es:

Daniel Minsky ; Thomas Grosjes

Co-director/es:

Claudio Abel Verraastro

Resumen

Las imágenes de Tomografía por Emisión de Positrones visualizan la actividad molecular en el tejido vivo. Para recuperar correctamente la distribución de la actividad es necesario conocer la anatomía del paciente y compensar los efectos de atenuación. Esto se realiza con un mapa de atenuación que se obtiene normalmente con otro tipo de imagen, como una Tomografía Computada. Recientemente se ha propuesto generar el mapa de atenuación usando métodos basados en imágenes, como las redes convolucionales. En este trabajo se explora la capacidad de las Redes Generativas Adversarias (GAN, del inglés) de crecimiento progresivo. Su exactitud es comparada comparada con GANs tradicionales de trabajos previos. A pesar de ser exitosas para la generación de imágenes de alta calidad 2D y poseer un entrenamiento más estable, su exactitud es menor a la obtenido en trabajos previos, alcanzando un Error Medio Promedio de 132+-22 HU contra 103+-18 HU obtenido en trabajos previos.

Palabras clave: Red Adveraria de Crecimiento Progresivo, Tomografía por Emisión de Positrones, Corrección de atenuación

Abstract

The Positron Emission Tomography images visualize molecular activity in living tissue. To correctly recover the activity distribution, the patient anatomical information is required to compensate the attenuation effects. This attenuation map is normally obtained through an external imaging technique, such as the Computed Tomography. Recently it has been proposed to recover the attenuation map using image-based methods, such as convolutional neural networks. In this work the capacity of progressive growing Generative Adversarial Networks (GANs) is explored. Their accuracy is compared to traditional GANs from previous works. Despite of being successful in high quality 2D image generation and posses more stable training, their accuracy falls behind prior works, achieving a Mean Average Error of 132+-22 HU against 103+-18 HU from previous results.

Keywords: Progressive Growing GAN, Generative Adversarial Networks, Positron Emission Tomography, Attenuation Correction

Introducción

El Tomógrafo por Emisión de Positrones (PET del inglés Positron Emission Tomography) en conjunto con el radio trazador 18 F-fluoro-2-deoxy-D-glucosa (FDG) se utiliza para obtener una imagen de metabolismo celular. La imagen de actividad celular es reconstruida a partir de la captura de fotones colineales que son generados por una aniquilación positrón-electrón. En su viaje desde su punto de emisión hasta uno de los detectores del tomógrafo PET, el fotón puede interactuar con la materia que atraviesa. En estas interacciones su energía disminuye y su trayectoria se modifica. Esta disminución provoca que el fotón sea descartado por el sistema de detección, ya que su trayectoria fue alterada y su información no es útil para la reconstrucción de la imagen de actividad. Entonces, el efecto de atenuación de la energía del fotón se traduce en una menor cantidad de eventos detectados. Esta disminución es más notoria en aquellas zonas donde la materia es más densa. Como resultado final, si no se tiene en cuenta este efecto las zonas del cuerpo con menor atenuación, como los pulmones y la piel, poseen un realce de actividad que no se condice con la actividad celular. Para corregir este efecto el tomógrafo PET debe ser equipado con otra tecnología que le permita sensar la geometría de las estructuras atenuantes, como lo es el Tomógrafo Computado (TC). Sin embargo el uso del CT no es inocuo para el paciente. Un estudio PET supone una dosis de entre 10 y 18 mSv de actividad (Ghotbi, 2007)(Kaushik, 2015), considerada una dosis “sustancial” por la Comisión Internacional de Protección Radiológica. El CT representa entre 1,3 y 8,0 mSv del total de radiación (siempre que se utilice un escaneo CT de baja dosis) (Ghotbi, 2007)(Kaushik, 2015)(Jallow, 2016). Por estos motivos la obtención de una imagen de atenuación sin necesidad de utilizar un CT es un paso de suma importancia para reducir los riesgos de la radiación durante un escaneo PET.

Varios métodos han sido probados para evitar el uso del CT, como prescindir de la imagen de atenuación utilizando modelos matemáticos que tengan en cuenta el efecto de la atenuación (Wu, 2005)(Pang, 2019) o generar los mapas de atenuación utilizando como base otro tipo de estudio, como por ejemplo la imagen por resonancia magnética (Armanious, 2020)(Bauer, 2019) o la imagen de PET sin corregir (NAC-PET del inglés Non Attenuation Corrected PET)(Liu, 2018)(Dong, 2019)(Colmeiro, 2020). Dentro de estas variantes la última, la generación de mapas de atenuación a partir de la imagen NAC-PET es la más interesante. Esta conversión no requiere ningún tipo de imagen base adicional y produce mapas de calidad con buena capacidad de corrección (Dong, 2020).

En este trabajo proponemos analizar el uso de Redes Generativas Adversarias 3D de Crecimiento Progresivo (ProGAN del inglés Progressive Growing Generative Adversarial Networks) en la tarea de convertir imágenes NAC-PET en CT, para luego ser utilizados como mapas de atenuación. Estas redes poseen un método de entrenamiento más estable que las redes tradicionales y han probado ser capaces de generar imágenes 2D de alta calidad (Karras, 2017). Este trabajo es comparado con métodos de generación estándar explorados con anterioridad (Colmeiro, 2020).

Los resultados y código para reproducir los resultados presentados se encuentran disponibles para la descarga en [https://github.com/RawthiL/PET_DeepAttCorrection].

Desarrollo

Las redes generativas adversarias de crecimiento progresivo se basan en el aumento progresivo de la resolución de la imagen generada. Propuestas originalmente para 2D (Karras, 2017), su extensión a 3D es relativamente directa. Conceptualmente, una GAN esta conformada por un par de redes neuronales que “compiten” entre si. Una red es denominada “Generadora” y su trabajo es tomar una entrada, que puede ser un vector aleatorio o un tipo de información y generar una salida que se parezca a una salida “real”. La otra red, denominada “Discriminadora” o “Crítico” toma una entrada y evalúa si la misma pertenece a la distribución de salida “real” o no. Durante el entrenamiento la red generadora es entrenada para generar imágenes que pertenezcan a la distribución objetivo mientras que la crítica es entrenada para detectar las imágenes generadas. El entrenamiento se continúa hasta que la red crítico no es capaz de detectar las imágenes falsas.

Normalmente un método externo es utilizado para evaluar la calidad de las imágenes generadas ya que el crítico puede no haber aprendido la distribución objetivo correctamente o el generador solo este generando un tipo de imagen (modo colapsado).

Dado el modo de entrenamiento de las GANs, son muy delicadas en su entrenamiento y generar imágenes de gran tamaño es difícil. Para ello se desarrollaron las redes progresivas, que construyen una imagen de alta calidad empezando por versiones más pequeñas y luego reutilizando lo aprendido para escalar el resultado. Esto es de especial interés para la generación de imágenes volumétricas ya que son de alta dimensionalidad.

Para el caso de la generación de mapas de atenuación para imágenes PET el generador toma como entrada una imagen volumétrica sin corregir (NAC-PET) y genera una imagen volumétrica que intenta imitar el CT (denominada CT sintético). Se trata de una red generativa condicional, en este caso condicionada a la información de la imagen NAC-PET. El crítico en este caso es una red que toma como entrada un par de imágenes volumétricas, la imagen de referencia (NAC-PET) y la imagen generada o real (CT real o CT sintético) y devuelve un valor numérico continuo que puede interpretarse como la "calificación" de la imagen. El concepto del generador ProGAN implementado se ilustra en la Figura 1 y del crítico en la Figura 2.

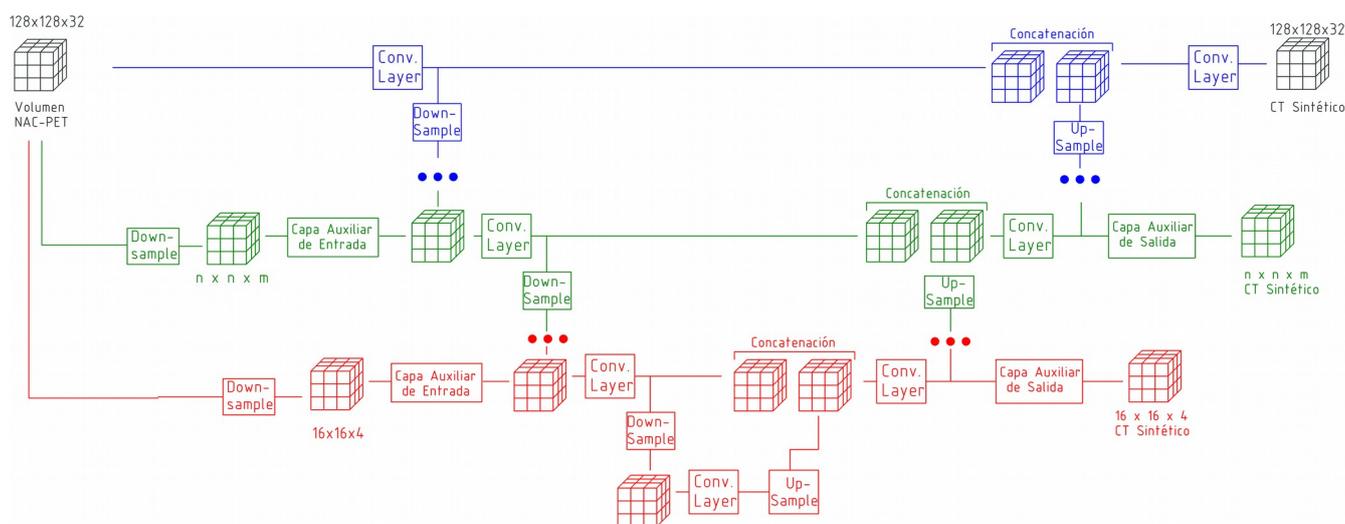


Figura 1: Esquema del generador ProGAN. Los bloques "Conv, Layer" representan un conjunto de operaciones convolucionales en tres dimensiones, los bloques "Down-" y "Up-Sample" son operaciones de disminución o aumento de resolución, los bloques "Capa Auxiliar" son solo utilizados durante el entrenamiento de la resolución marcada. Cada color representa una resolución. La red se inicia en la resolución mas baja (color rojo) y va aumentando progresivamente hasta la resolución más alta (color azul).

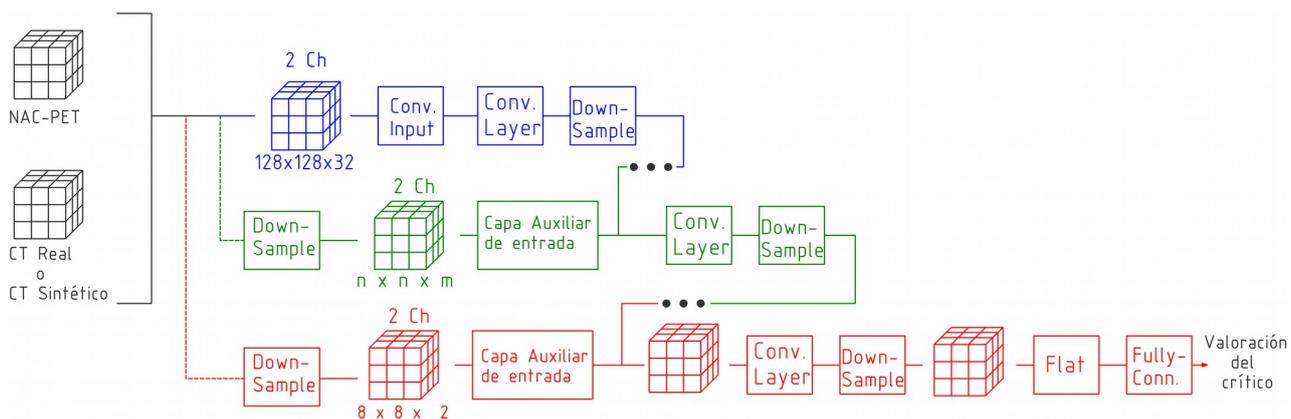


Figura 2: Esquema del crítico ProGAN. Los bloques "Conv, Layer" representan un conjunto de operaciones convolucionales en tres dimensiones, los bloques "Down-" y "Up-Sample" son operaciones de disminución o aumento de resolución, el bloque "Flat" es una operación de aplanamiento, el bloque "Fully Conn." es una capa totalmente conectada, los bloques "Capa Auxiliar" son solo utilizados durante el entrenamiento de la resolución

marcada. Cada color representa una resolución. La red se inicia en la resolución mas baja (color rojo) y va aumentando progresivamente hasta la resolución más alta (color azul).

El entrenamiento se realiza en etapas, comenzando por el nivel de resolución más bajo. Cada nivel es entrenado hasta que la calidad de las imágenes generadas se estabiliza. Luego se añade el siguiente nivel y se re-entrena el conjunto de redes. Este proceso se repite hasta alcanzar el nivel de resolución deseado.

Set de Datos: Para el entrenamiento, validación y prueba de la red se utilizó una serie de datos públicos provistos en el Archivo de Imágenes del Cáncer (TCIA, del inglés The Cancer Image Archive) (Clark, 2013). Los sets se construyeron utilizando los métodos de pre-procesamiento presentados por Colmeiro (2020). El set de datos de entrenamiento está compuesto de 108 pares de volúmenes NAC-PET + CT, de 62 pacientes distintos, provenientes del set de Carcinomas Espinocelulares de Cabeza y Cuello (HNSCC, del inglés Head and Neck Squamous-Cell Carcinoma) (Bejarano, 2018). El set de validación, también del set HNSCC, es de 10 muestras, de 7 pacientes distintos. El set de prueba está compuesto de imágenes de múltiples orígenes, todas encontradas en TCIA, resultando en 133 muestras de distintos pacientes.

Parámetros de la red neuronal: Dado que el tamaño de las imágenes volumétrica a generar es muy grande ($128 \times 128 \times 256 = 4194304$ vóxeles) se estableció como objetivo un volumen de $128 \times 128 \times 32$, que representa una porción axial del volumen a generar. Dada esta condición, la generación de un volumen completo se realiza con sucesivas aplicaciones de la red, concatenando los volúmenes resultantes.

Se eligió utilizar 5 niveles de resolución, siendo la resolución más pequeña $8 \times 8 \times 2$. El generador implementa 160, 80, 40, 20 y 10 canales, desde el nivel de resolución más pequeño al más grande. Cada nivel aplica dos capas de convolución 3D con un tamaño de filtro de $3 \times 3 \times 3$ vóxeles. Los pesos de los filtros están normalizados por un valor determinado por la escala de He (He, 2015). El generador no utiliza normalización por tandas, sino que se utiliza una normalización por característica, denominada "normalización de píxel" siguiendo los parámetros en Colmeiro (2020). El crítico también se compone de 5 niveles, utilizando 256, 128, 64, 32 y 16 canales respectivamente. Cada nivel de resolución implementa dos capas convolucionales con filtros de $3 \times 3 \times 3$ vóxeles. El crítico termina en una única neurona con activación lineal que entrega la calificación de la imagen de entrada. El crítico no aplica ni normalización por tandas ni por píxel. Las activaciones de las capas convolucionales, tanto del generador como el crítico, son unidades rectificadas (ReLU, del inglés Rectifier Linear Unit) con pérdida igual a 0,2.

Entrenamiento: El entrenamiento es realizado por niveles de resolución. Cada nivel es entrenado durante 1500, 3000, 6000, 8000 y 18000 iteraciones, desde el nivel más bajo al más alto respectivamente. Cada iteración consta de 5 pasos de entrenamiento del crítico y 1 paso de entrenamiento del generador. El generador fue entrenado utilizando un algoritmo de momentos adaptados (ADAM, del inglés Adaptive Moment estimation), con velocidad de entrenamiento igual a 0.0001, momento de primer orden igual a 0,0 y de segundo orden igual a 0,99. El crítico fue entrenado con un algoritmo de propagación de raíz de la media cuadrada (RMSProp, del inglés Root Mean Square Propagation), con velocidad de entrenamiento igual a 0.0005. El entrenamiento se realizó utilizando el entorno de trabajo TensorFlow (Abadi, 2016).

Resultados

La red resultante, denominada ProGAN, fue probada comparada con la topología presentada por Colmeiro (2020) (SplitGradGAN) y una red de las mismas características pero entrenada sin entrenamiento progresivo (GAN Base). Para ellos se utilizaron las métricas de relación de señal a ruido pico (PSNR, del inglés Peak Signal to Noise Ratio), error medio absoluto (MAE, del inglés Mean Absolute Error) y correlación cruzada normalizada (NCC, del inglés Normalized Cross Correlation). Los resultados están recopilados en la tabla 1. Se puede observar que la red presenta una mejora con respecto al método GAN Base, pero no llega a los valores de exactitud observados en la SplitGradGAN.

La diferencia de calidades de imagen se evidencia en la Figura 3, donde se comparan las reconstrucciones de los métodos en una muestra del set de prueba. Se observa que a pesar de

existir una mejora los resultados con respecto a la GAN Base, la ProGAN no presenta una buena correlación con la imagen objetivo, como si se observa en la SplitGradGAN.

Topología	Set de Datos	PSNR [db]	MAE [HU]	NCC[-]
GAN Base	Validación	17,5±1,8	147±38	0,54±0,14
	Testeo	17,0±1,4	143±25	0,57±0,08
ProGAN	Validación	18,8±1,3	115±18	0,65±0,09
	Testeo	17,4±1,3	132±22	0,58±0,07
SplitGradGAN	Validación	19,9±1,3	89±10	0,76±0,05
	Testeo	18,6±1,5	103±18	0,72±0,06

Tabla 1: Resultados de validación (10 muestras) y de testeo (133 muestras) para las distintas topologías estudiadas. La topología presentada se encuentra resaltada.

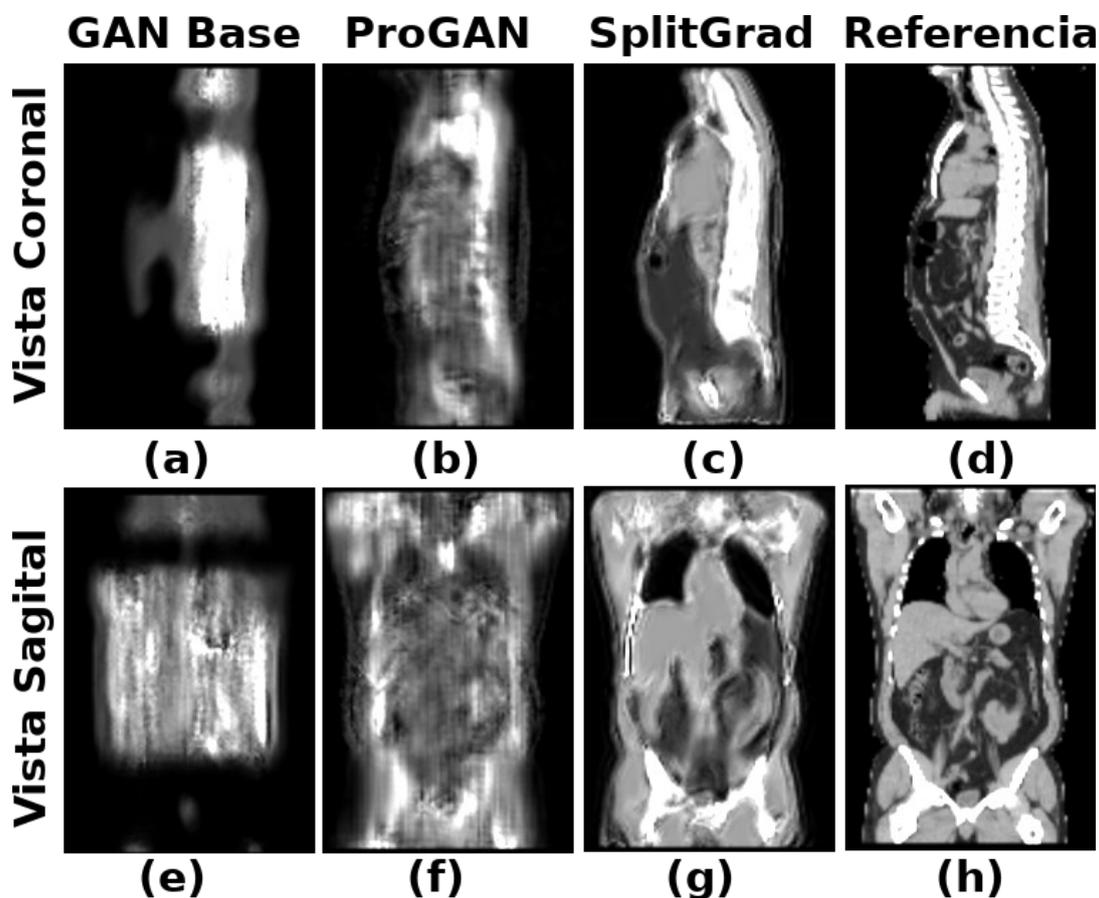


Figura 3: Vistas coronales (a,b,c,d) y sagitales (e,f,g,h) de los CT sintéticos generados por la GAN Base (a,e), por la ProGAN (b,f), por la SplitGradGAN (c,h) y la imagen de referencia (d,h).

Como era esperable se observa una mejora con respecto al método GAN Base, pero se puede concluir que la mera adición de una estrategia de entrenamiento escalonado no compensa la dificultad de generar imágenes 3D de alta calidad. Se continuará explorando los modelos generativos, prestando atención a los modelos generativos con espacios de muestreo fijo, que regularizan la generación de las imágenes 3D, como los presentados en (Zhu, 2018).

Referencias

- Ghotbi, N., Iwanaga, M., Ohtsuru, A., Ogawa, Y., & Yamashita, S. (2007). Cancer screening with whole-body PET/CT for healthy asymptomatic people in Japan: re-evaluation of its test validity and radiation exposure. *Asian Pacific Journal of Cancer Prevention*, 8(1), 93.
- Kaushik, A., Jaimini, A., Tripathi, M., D'Souza, M., Sharma, R., Mondal, A., ... & Dwarakanath, B. S. (2015). Estimation of radiation dose to patients from 18FDG whole body PET/CT investigations using dynamic PET scan protocol. *The Indian journal of medical research*, 142(6), 721.
- Jallow, N., Christian, P., Sunderland, J., Graham, M., Hoffman, J. M., & Nye, J. A. (2016). Diagnostic reference levels of CT radiation dose in whole-body PET/CT. *Journal of Nuclear Medicine*, 57(2), 238-241.
- Wu, T. H., Chu, T. C., Huang, Y. H., Chen, L. K., Mok, S. P., Lee, J. K., ... & Lee, J. J. (2005). A positron emission tomography/computed tomography (PET/CT) acquisition protocol for CT radiation dose optimization. *Nuclear medicine communications*, 26(4), 323-330.
- Pang, L., Zhu, W., Dong, Y., Lv, Y., & Shi, H. (2019). Zero-Extra-Dose PET Delayed Imaging with Data-Driven Attenuation Correction Estimation. *Molecular Imaging and Biology*, 21(1), 149-158.
- Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., ... & Yang, B. (2020). MedGAN: Medical image translation using GANs. *Computerized Medical Imaging and Graphics*, 79, 101684.
- Bauer, D. F., Schnurr, A. K., Russ, T., Goertler, S., Schad, L. R., Zoellner, F. G., & Chung, K. (2019, April). Synthesis of CT Images Using CycleGANs: Enhancement of Anatomical Accuracy. In *International Conference on Medical Imaging with Deep Learning--Extended Abstract Track*.
- Liu, F., Jang, H., Kijowski, R., Zhao, G., Bradshaw, T., & McMillan, A. B. (2018). A deep learning approach for 18 F-FDG PET attenuation correction. *EJNMMI physics*, 5(1), 1-15.
- Dong, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Curran, W. J., ... & Yang, X. (2019). Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. *Physics in Medicine & Biology*, 64(21), 215016.
- Colmeiro, R. R., Verrastro, C., Minsky, D., & Grosjes, T. (2020). Whole Body Positron Emission Tomography Attenuation Correction Map Synthesizing using 3D Deep Generative Adversarial Networks.
- Dong, X., Lei, Y., Wang, T., Higgins, K., Liu, T., Curran, W. J., ... & Yang, X. (2020). Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. *Physics in Medicine & Biology*, 65(5), 055011.
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., ... & Tarbox, L. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6), 1045-1057.
- Bejarano, T., De Ornelas Couto, M., & Mihaylov, I. Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment dataset. the cancer imaging archive; 2018.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- Zhu, J. Y., Zhang, Z., Zhang, C., Wu, J., Torralba, A., Tenenbaum, J., & Freeman, B. (2018). Visual object networks: Image generation with disentangled 3D representations. In *Advances in neural information processing systems* (pp. 118-129)