 [Traduzido do Inglês para o Português - www.onlinedoctranslator.com](https://www.onlinedoctranslator.com/pt/?utm_source=onlinedoctranslator&utm_medium=docx&utm_campaign=attribution)

***Anais do ICSP2022***

GST: um transmissor de sinal gráfico inspirado no cérebro para segmentação de imagens biomédicas

Caiqing Jian 1, Yongbin Qin1\*, Lihui Wang1, Hui Xia 2, Yuemin Zhu3

1. Laboratório Chave de Análise de Imagens Médicas Inteligentes e Diagnóstico Preciso da Província de Guizhou, Laboratório Chave Estatal de Big Data Público,

Faculdade de Ciência e Tecnologia da Computação, Universidade de Guizhou, Guiyang, China.

1. Departamento de Cirurgia Cardiotorácica, Quarto Centro Médico, PLA General Hospital, Pequim 100048, China.
2. Univ Lyon, INSA Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, França.

E-mail: gs .cqjian21@gzu.edu.cn , Autor correspondente: wlh1984@gmail.com ,

117 34048@zju.edu.cn , xiahui304@163.com , zhu@creatis.insa-lyon.fr

Resumo—A segmentação automática de imagens biomédicas é de grande importância para o diagnóstico auxiliado por computador. Diferente dos modelos de segmentação existentes baseados em rede neural convolucional ou mecanismo de atenção com modo de previsão pixel a pixel, este artigo propõe um novo transmissor de sinal de gráfico (GST) que regride a probabilidade de patch diretamente do nó semântico e recursos de borda evoluídos. As imagens passam por um codificador para derivar as características semânticas, cada voxel do qual é tomado como um nó, e o vetor de característica correspondente ao longo da dimensão do canal é tomado como a característica do nó da camada gráfica inicial, e as posições relativas entre diferentes nós são considerados como os recursos de borda. Com o processo feedforward, os recursos de nó e borda das camadas seguintes são evoluídos pela fusão dos recursos dos nós de origem, nós e arestas de destino. As características do nó das últimas camadas do gráfico são mapeadas diretamente de volta para o patch no espaço semântico com o módulo de dobra e geram o mapa de segmentação. Nas tarefas de segmentação semântica e de instância para o conjunto de dados da glândula, experimentos comparativos mostram que o modelo proposto supera os métodos de última geração (SOTA), e experimentos de ablação demonstram a eficácia do GST.

Palavras-chave — gráfico, inspirado no cérebro, transmissão de sinal, imagem biomédica, segmentação, aprendizagem profunda.

# EU. EUNTRODUÇÃO

A segmentação é um problema desafiador no processamento de imagens médicas, geralmente é implementado manualmente por médicos experientes, o que não é apenas demorado e trabalhoso, mas também introduz o problema de padrões de rotulagem inconsistentes. Nos últimos anos, o aprendizado profundo alcançou desempenho de ponta em muitas aplicações na segmentação de imagens médicas, o que deve resolver os problemas acima.

|  |
| --- |
| **978-1-6654-6056-9/22/$31,00 ©2022 IEEE** |

Como um marco para a segmentação semântica, a rede totalmente convolucional (FCN) [2] permite que a CNN seja naturalmente adaptada para tarefas de previsão densas pixel a pixel. Com base no FCN, vários modelos aprimorados foram propostos. Por exemplo, os modelos da família DeepLab [3-5] consideram informações contextuais na segmentação semântica por meio de convolução dilatada, PSPNet [6] combina recursos multiescala para melhorar o desempenho do modelo, UNet e suas variantes [7-9] introduzem conexões de salto entre codificador e decodificador para promover os detalhes de segmentação e se tornaram modelos de referência para segmentação de imagens médicas.

2022 16

ª Conferência Internacional IEEE sobre Processamento de Sinais (ICSP) | 978-1-6654-6056-9/22/$31,00 ©2022 IEEE | DOI

: 10.1109/ICSP56322.2022.9965289

Com o sucesso do Transformer [10], mecanismos de atenção foram gradativamente sendo aplicados em tarefas de visão. Por exemplo, Unet não local [11] e TransUNet [12] aplicam mecanismos de atenção na arquitetura codificador-decodificador para capturar semântica contextual e dependências de longo alcance. No entanto, simplesmente introduzir o módulo de auto-atenção aumentará a complexidade computacional excessiva. Para resolver este problema, Swin-Transformer [13] usa uma arquitetura hierárquica para calcular a atenção através de janelas móveis. O Swin-UNet [14] combina as vantagens do UNet e do SwinTransformer, no entanto, não funciona bem em conjuntos de dados de pequena escala, pois não possui o viés indutivo da convolução. Para resolver esse problema, o UTNet [15] combina autoatenção com convolução, que pode explorar totalmente o viés indutivo de redes convolucionais e as capacidades de modelagem de longo alcance dos mecanismos de atenção. Ao mesmo tempo, a complexidade computacional é reduzida diminuindo a resolução dos vetores chave e valor. Desde que o ViT [16] foi proposto, os modelos baseados em transformadores atingiram gradualmente o desempenho do estado da arte (SOTA) para várias tarefas visuais. No entanto, o ConvNeXt subsequente [17] e o RepLKNet [18] também podem obter desempenho comparável ao ViT ao usar kernels de convolução de grande profundidade e seguir a arquitetura e a estratégia de treinamento do ViT.

Nos modelos mencionados acima, a imagem é considerada como uma matriz com estrutura de grade regular. No entanto, em certos casos, expressar os dados da imagem com um gráfico obterá melhor desempenho. Hanzhe [19] et al. proposta de convolução dinâmica de grafos intraclasse para evitar interferência causada por atributos de categorias irrelevantes; Yanda [20] et al. propôs explorar completamente a associação entre região e limite usando rede neural de gráfico (GNN) para melhorar o desempenho da segmentação semântica. No entanto, os principais problemas das GNNs existentes são: (1) falta de estratégias efetivas de codificação de posição relativa, (2) a associação entre os nós é temporal e sua evolução com o processo feedforward não é armazenada, (3) as informações transmitidas do nó nó de origem para o nó de destino depende principalmente do estado do nó de origem, enquanto o estado do nó de destino e da borda não são considerados. Para lidar com essas questões, uma rede de transmissão de sinal gráfico (GST) inspirada no cérebro é proposta neste trabalho. No GST, os neurônios cerebrais são representados por nós e as sinapses são expressas como arestas. Assumimos que o sinal transmitido do nó de origem ao nó de destino deve conter os três seguintes elementos: o que, onde e intensidade do sinal, enquanto satisfaz o princípio da transmissão de sinal, ou seja, a transmissão do sinal entre os nós depende não apenas do seu próprio estado, mas também o estado da borda. Com base nessa suposição, projetamos uma rede de grafos que transmite as características do nó e da borda, bem como as posições relativas do nó para a próxima camada através de uma porta para controlar a intensidade do sinal. Na seção seguinte,

|  |
| --- |
| Fig. 1. O pipeline do GST proposto para segmentação de imagens biomédicas |

# II. MÉTODOS

O pipeline do GST proposto é mostrado na Fig.1, a imagem é dividida em vários patches. Para simplificar, apenas patches 4 × 4 não sobrepostos são mostrados aqui, o método de divisão real é detalhado posteriormente. O GST consiste em quatro etapas. Passo 1: Extraia características semânticas de uma imagem com codificador de backbone de convolução e obtenha os nós semânticos correspondentes aos patches; Etapa 2: atribuir nós semânticos ao gráfico e apará-lo de acordo com a correlação de recursos entre os nós; Etapa 3: O mecanismo de transmissão de sinal é usado para comunicar entre os nós de origem (azul) e o nó de destino (vermelho), e as informações agregadas de vários nós de origem são transformadas por uma rede perceptual multicamada (MLP) no nó de destino; Passo 4: O vetor de característica do nó gerado na etapa 3 é mapeado para o vetor de probabilidade de classe do patch correspondente usando uma camada MLP, então vários patches são unidos no mapa de probabilidade de classe de toda a imagem e, finalmente, arredondados para o mapa de rótulo previsto. Os detalhes são descritos nas subseções a seguir.

## A. Construção de nós semânticos

Todo o codificador e a camada do decodificador de nível inferior na UNet [7] são usados ​​para extrair as características semânticas dos nós do grafo. Dada a imagem de entrada XR128 192 3  , o recurso semântico resultante é E(X) R16 24 512  . Total de 16×24=384 nós semânticos são obtidos, cada nó corresponde a um patch com tamanho de 8 × 8 na Fig. 1, e a dimensão do recurso do nó é 512. Para evitar efeitos de limite causados ​​por patches não sobrepostos, além do divisão de manchas não sobrepostas (Fig.2(a)), três divisões de manchas sobrepostas também são usadas neste trabalho (Fig.2(b)-(d)). Um total de 1617 (384+425+400+408=1617) nós semânticos (patches) são obtidos para cada imagem X. As características dos nós são anotadas como HR1617 512 , e a probabilidade de classe de patches sobrepostos é calculada em média durante a previsão.

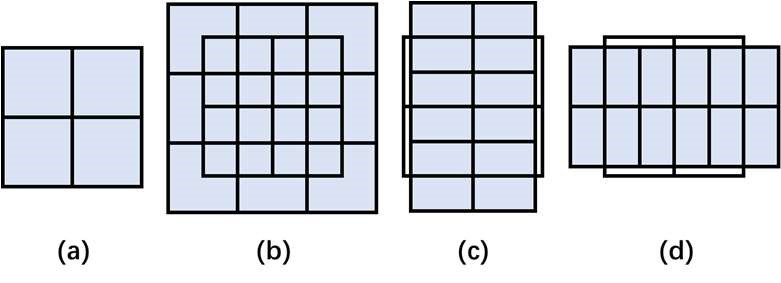


Fig. 2. Diagrama esquemático da amostragem de fragmentos sobrepostos

## B. Amostragem da estrutura do gráfico

Para reduzir a complexidade computacional e evitar a interferência de informações de nós irrelevantes, a associação rij entre o nó de destino i e o nó de origem j é calculada

com base nos recursos do nó hi e hj ,

rij  hihj. (1)

Apenas os nós que são mais relevantes para o nó de destino (rij > threshold) são mantidos. Conforme mostrado na Fig. 3, o nó vermelho representa o nó de destino, apenas três nós de origem mais relevantes (azuis) à esquerda são mantidos após a amostragem do gráfico.

Fig. 3. Amostragem da estrutura do gráfico

Amostra

## C. Transmissão de Sinal e Mecanismo de Transformação

A transmissão do sinal e o mecanismo de transformação do GST são mostrados na Fig. 4. O nó fonte (Src) e o nó destino (Dst) são inicializados com HR1617 512 . A aresta entre o nó i e o nó j na primeira camada (eij0 ) é inicializada com sua posição relativa rposij , definida pela projeção linear da coordenada relativa coordij e distância distij entre os nós i e j em um sistema de coordenadas cartesianas estendidas por nós.

rposeu j proj dist eu jcoordeu j, (2)

|  |
| --- |
| Fig. 4. Transmissão de sinal e mecanismo de transformação de GST |

onde || indica a operação de concatenação. O recurso do nó de origem hjl-1, o recurso do nó de destino hil-1 e seu recurso de borda eijl-1 da l-1ª camada são fundidos primeiro por meio da concatenação e depois transformados com duas camadas MLP para gerar o sinal transmitidoijl e o recurso de borda eijl da lª camada, respectivamente,

|  |  |
| --- | --- |
| eu jeuWheueul1eijl1heuj1, | (3) |
| eijeuO que eleeueul1eijl1heuj1, | (4) |

onde Wl e Wel representam os parâmetros das camadas MLP.

Este sinal transmitido ijl passa por uma porta para determinar quanto de recursos de borda pode ser transmitido para o nó de destino na lª camada, o coeficiente de porta gijl é formulado como

exp(eu)eu eu j

geu jeu. (5)máx{exp(eu j)}

jeu

Assim, o sinal de borda efetivo transmitido ao nó de destino da lª camada é escrito como

e'lijN1eu(geij ijl eu) (6)

jeu

onde i é a vizinhança do nó i. Neste trabalho, a vizinhança é limitada na região com raio de 8. Para descrever com precisão o sinal de borda transmitido da camada anterior, dividimos o sinal de borda em K grupos, os sinais em cada grupo são processados ​​com as equações (3) ~(6), e o recurso do nó de destino i na lª camada é finalmente derivado por

k 

h Wi  ne'ijk hi (7)eueul1

k=1 

onde e'lijk é o sinal de borda efetivo transmitido para a l-ésima camada com o k-ésimo grupo de sinais, Wnl significa os parâmetros da camada MLP.

## D. Segmentação no Espaço Semântico

Propomos usar a operação de dobra para mapear diretamente o vetor de probabilidade de classe de um patch regredido por MLP no espaço semântico para gerar um mapa de probabilidade bidimensional de acordo com as posições dos pixels. A Fig.5 ilustra o processo de operação de dobra, na qual um vetor de probabilidade de classe de 64 dimensões de um determinado nó é dobrado como um patch 8×8. A dobragem de todos os vetores de probabilidade de nós resulta no mapa de probabilidade de classe final.

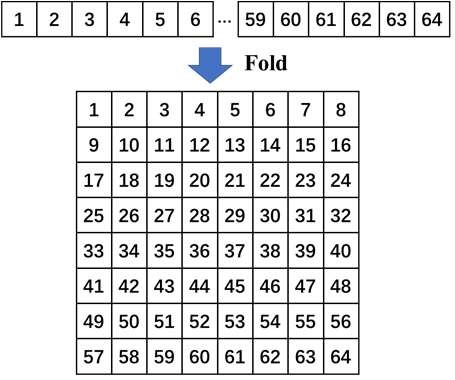


Fig. 5. Correspondência entre elementos de um vetor de probabilidade de classe de nó e pixels de um patch.

# III. ECONFIGURAÇÕES EXPERIMENTAIS

## A. Descrição de dados e pré-processamentos

O conjunto de dados usado neste trabalho é do GlaS Challenge Contest [1], que possui 165 imagens de histologia coradas com H&E e rótulos de instância correspondentes. O organizador o dividiu em três subconjuntos, a saber, conjunto de treinamento, conjunto de teste A e conjunto de teste B, que podem ser usados ​​para avaliar a segmentação semântica e o desempenho da segmentação de instância do modelo. O tamanho da imagem original é de cerca de 522 × 775, os experimentos de treinamento e teste são realizados com imagens redimensionadas para 128 × 192, exceto dois experimentos comparativos que requerem o tamanho fixo da imagem de entrada (Swin-UNet 224 × 224, UTNet 256 × 256). Observe que, ao calcular métricas quantitativas para avaliar o desempenho de segmentação de diferentes métodos, o mapa de rótulos de saída é redimensionado para 256 × 384.

## B. Hiperparâmetros Experimentais e Estratégia de Treinamento

Usamos a estrutura Pytorch para treinar e testar em um único

GPU 2080Ti. O otimizador é Adam, taxa de aprendizado inicial lr = 0,001, BatchSize = 1, Epoch = 100. As funções de perda são MSE e IoU. A estratégia de decaimento da taxa de aprendizado é ReduceLROnPlateau em que paciência = 3 e fator = 0,8 . Além disso, métodos de aumento de dados, incluindo inversão vertical e horizontal aleatória, rotação aleatória (intervalo: −30◦ a 30◦), ​​deformação aleatória, cor aleatória e transformação de brilho, desfoque gaussiano aleatório são usados.

# 4. RESULTADOS

## A. Resultados experimentais comparativos

Para verificar o desempenho do GST proposto, ele foi comparado com vários modelos SOTA, incluindo DeepLabV3 [4], UNet [7], UNet++ [8], UNet3+ [9], SwinUNet [14], UTNet [15]. Também implementamos duas redes de atenção em grafos como experimentos comparativos sobre o mecanismo de passagem de mensagens em grafos, a saber, GAT [22] e o transformador de vizinhos de grafos (GNT), onde o GNT usa o Transformer's

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MESA EU  TELEQUANTITATIVOEAVALIAÇÃOMÉTRICAS PARAGTERRASEGMENTAÇÃOCITHDDIFERENTEMÉTODOS   |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | modelos | IoU | Obj\_Dice: Total/A/B | | | | F1: Total/A/B | | | | HD: Total/A/B | | | Parâmetros | | GST | 0,8721 | 0,8901 | 0,9084 | 0,8353 | 0,8718 | | 0,8985 | 0,7916 | 29.41 | | 22.60 | 49,84 | 17,29M | | GAT | 0,8498 | 0,8395 | 0,8503 | 0,8068 | 0,7764 | | 0,8073 | 0,6837 | 45,89 | | 38.05 | 69,41 | 17,31M | | GNT | 0,8524 | 0,8466 | 0,8610 | 0,8036 | 0,8003 | | 0,8240 | 0,7294 | 42.18 | | 35.11 | 63,41 | 17,32M | | DeepLabV3 | 0,8091 | 0,7270 | 0,7336 | 0,7071 | 0,6434 | | 0,6598 | 0,5944 | 82,82 | | 75,43 | 104,9 | 15,34M | | SwinUNet | 0,8300 | 0,7529 | 0,7573 | 0,7399 | 0,6999 | | 0,7171 | 0,6483 | 72,73 | | 69,34 | 82,88 | 41,96M | | UNet | 0,8202 | 0,7881 | 0,7989 | 0,7558 | 0,7017 | | 0,7355 | 0,6002 | 59,84 | | 53,93 | 77,55 | 17,27M | | UNet++ | 0,8235 | 0,8034 | 0,8216 | 0,7486 | 0,7233 | | 0,7473 | 0,6515 | 58,72 | | 49,41 | 86,66 | 47,18M | | UNet3+ | 0,8325 | 0,8209 | 0,8376 | 0,7707 | 0,7446 | | 0,7753 | 0,6526 | 50.01 | | 40,91 | 77,31 | 26,98M | | UTNet | 0,8557 | 0,8288 | 0,8411 | 0,7920 | 0,7823 | | 0,8038 | 0,7177 | 47.07 | | 41,89 | 62,61 | 10,01M | |  |  |  |  |  | MESA EU | | EU |  |  | |  |  |  | |  |  | TELEQUANTITATIVORRESULTADOS DE DIFERENTES EXPERIMENTOS DE ABLAÇÃO | | | | | | | | | |  |  | | modelos | IoU | Obj\_Dice: Total/A/B | | | | F1: Total/A/B | | | | HD: Total/A/B | | | Parâmetros | | F(T(E(X))) | 0,8721 | 0,8901 | 0,9084 | 0,8353 | 0,8718 | | 0,8985 | 0,7916 | 29.41 | | 22.60 | 49,84 | 17,29M | | F(L(E(X))) | 0,8485 | 0,8371 | 0,8486 | 0,8028 | 0,7836 | | 0,8125 | 0,6969 | 45,63 | | 38,71 | 66,41 | 13,91 milhões | | F(G(E(X))) | 0,8462 | 0,8353 | 0,8465 | 0,8018 | 0,7817 | | 0,8189 | 0,6701 | 46.16 | | 39,90 | 64,93 | 11,27M | | F(E(X)) | 0,8456 | 0,8334 | 0,8398 | 0,8144 | 0,7773 | | 0,8002 | 0,7088 | 48.15 | | 42.17 | 66.08 | 6,27M | | D(E(X)) | 0,8362 | 0,7868 | 0,7934 | 0,7671 | 0,7258 | | 0,7518 | 0,6474 | 60,64 | | 55.23 | 76,88 | 6,24M | | U(E(X)) | 0,8286 | 0,7844 | 0,7872 | 0,7759 | 0,7385 | | 0,7519 | 0,6982 | 61,43 | | 55,30 | 79,84 | 6,23M | |

codificador [23] para passagem de mensagem em gráfico. Usamos IoU para avaliar o desempenho da segmentação semântica. Quanto à segmentação da instância, três métricas reconhecidas são usadas [1], incluindo a pontuação Dice em nível de objeto (Obj\_Dice), que mede o coeficiente Dice de segmentação em nível de instância, pontuação F1 (F1), que mede a precisão da detecção de instâncias de glândula, e Distância Hausdorff (HD) em nível de objeto, que indica a semelhança entre o limite previsto da glândula e aquele do rótulo. Os resultados quantitativos são apresentados na Tabela I, na qual três métricas para segmentação de instância são avaliadas no conjunto de teste total (Total), conjunto de teste A (A) e conjunto de teste B (B) , respectivamente. Descobrimos que o GST supera os modelos comparativos em métricas semânticas e de instância. Comparado com DeepLabV3, SwinUNet, UNet, UNet++, UNet3+, UTNet, GAT e GNT, o GST é melhorado em 7,8%,

2,6% e 2,3% respectivamente em IoU; 22,4%, 18,2%, 12,9%,

10,8%, 8,4%, 7,4%, 6,0% e 5,1% em Obj\_Dice; 35,5%, 24,6%, 24,2%, 20,5%, 17,1%, 11,4%, 12,3% e 8,9% na pontuação F1. A distância de Hausdorff é reduzida para 29,41, quanto menor o HD, mais próximo o limite de segmentação está do rótulo. Esses resultados validam a superioridade do GST na segmentação da glândula.

## B. Estudo de ablação

Para verificar a eficácia de nossos módulos propostos, as ablações de GST incluem: (1) Remoção de GST; (2) Mantendo a arquitetura geral e substituindo GST por GCN [21] e convoluções em profundidade de tamanho grande, ambas com o mesmo campo receptivo que GST; (3) Substituindo o módulo Fold por upsampling e convolução transposta. Para conveniência de comparação, denotamos o Encoder, GST, Fold e GCN como E, T, F, G, respectivamente, e a convolução em profundidade de tamanho grande como L, a convolução transposta e upsampling como D e U. Os resultados quantitativos das experiências de ablação são apresentados na Tabela II. As três primeiras linhas demonstram que o GST tem capacidade de modelagem de contexto mais forte do que kernels de convolução em profundidade de tamanho grande e GCN. Comparado com F(E(X)), F(T(E(X))) tem uma melhoria de 3,1% em IoU, 6,8% em Obj\_Dice, 12. 2% na pontuação F1，e atinge um HD 18,7% menor. Estes resultados demonstram a eficácia do GST. As últimas três linhas de resultados de ablação mostram a superioridade do módulo de dobra sobre a convolução transposta e o upsampling.

# v. DISCUSSÃO

|  |
| --- |
| Fig. 6. Visualização dos resultados da segmentação da instância da mesma amostra no conjunto de teste por diferentes modelos. |

À medida que a imagem passa por várias camadas convolucionais do codificador, as informações transmitidas mudam de detalhes para semântica. FCN [2] projeta o vetor de características no espaço semântico para o vetor de probabilidade de classe e, em seguida, usa a convolução transposta para obter um mapa de probabilidade do mesmo tamanho da imagem original, mas o upsampling levará à perda dos detalhes dos limites da instância. Conforme mostrado pelo retângulo vermelho de D(E(X)) e DeepLabV3 na Fig. 6, a perda de detalhes faz com que as duas instâncias fiquem juntas. Para lidar com esse problema, os modelos das UNets restauram informações detalhadas introduzindo recursos rasos. Pode-se observar que a família UNet possui um alto grau de discriminação de instâncias. UNet3+ atinge uma métrica de segmentação de instância de (Obj\_Dice: 0,8209, F1: 0,7446, HD: 50,01), que é significativamente melhor do que as métricas (Obj\_Dice: 0,7868, F1: 0,7258, HD: 60,64) de D(E(X)). Mesmo a métrica de segmentação de instância da UNet é geralmente maior do que a do modelo com convolução transposta.

No entanto, acreditamos que recursos superficiais podem facilmente introduzir informações semanticamente irrelevantes, o que interfere no treinamento e na previsão do modelo. Por exemplo, as métricas de IoU e instância da família UNet são menores que a de F(E(X)) que realiza a segmentação diretamente no espaço semântico. Além disso, devido à diversidade de características da imagem histopatológica e ao formato irregular das glândulas, informações semanticamente irrelevantes podem facilmente levar a um falso positivo na UNet3+ ou falso negativo na UNet.

Também descobrimos que aumentar o campo receptivo pode tornar os limites mais completos e suaves, como os de GST, GAT, GNT, F(L(E(X))), F(G(E(X))) e UTNet . Por outro lado, os limites de modelos com campo receptivo pequeno, como UNet e Swin-UNet, são mais fragmentados. No entanto, simplesmente aumentar o campo receptivo facilmente fará com que instâncias com pequenas áreas sejam perdidas, como as caixas vermelhas mostradas no canto inferior esquerdo de GAT, GNT, F(L(E(X))), F(G(E (X))) e UTNet na Fig.6. Comparado com grandes núcleos de convolução ou autoatenção, o mecanismo de transmissão de sinal do GST pode modelar melhor a semântica do contexto. Além disso, a amostragem da estrutura do grafo do GST faz com que o nó de destino preste atenção aos nós de origem importantes com correlações mais altas, evitando assim a detecção perdida de pequenas instâncias.

# VI. CONCLUSÃO

Este artigo propõe uma nova rede de transmissão de sinal de gráfico GST, inspirada no cérebro, que considera totalmente o princípio da transmissão de sinal: a transmissão de sinal entre os nós depende não apenas do estado dos nós, mas também do estado de sua borda. A codificação de posição relativa do nó de origem e do nó de destino é introduzida na borda e o portão é usado para gerar o coeficiente de gating, de modo que a mensagem que passa no gráfico satisfaça os três elementos de quê, onde e intensidade do sinal, o que melhora muito a rede no aprendizado de informações semânticas contextuais. Em seguida, o Grafo é amostrado com base na correlação, o que não apenas reduz a complexidade computacional, mas também evita a interferência de informações de nós irrelevantes. Por fim, os recursos do nó são mapeados diretamente de volta ao patch no espaço semântico, que não apenas preserva as informações de localização espacial, mas também melhora a generalização do modelo. Tanto os experimentos comparativos quanto os experimentos de ablação mostram que o GST tem desempenho de segmentação significativamente melhor do que os modelos existentes e tem vantagens potenciais de aplicação em outras tarefas de segmentação de imagens médicas.

# RECONHECIMENTO

Este trabalho foi apoiado pelas Fundações Nacionais de Ciências da Natureza da China (Concessão No. 62161004), o Projeto Chave do Fundo de Ciência e Tecnologia de Guizhou (QianKeHe ZK [2021] Chave 002), Fundações de Ciências da Natureza da Província de Guizhou (QianKeHe [2020] 1Y255 ).

# REFERÊNCIAS

1. Sirinukunwattana, Korsuk, et al. "Segmentação de glândulas em imagens de histologia do cólon: o concurso de desafio de glas." Análise de imagens médicas 35 (2017): 489-502.

1. Long, Jonathan, Evan Shelhamer e Trevor Darrell. "Redes totalmente convolucionais para segmentação semântica." Anais da conferência IEEE sobre visão computacional e reconhecimento de padrões. 2015.
2. Chen, Liang-Chieh, et al. "Deeplab: segmentação de imagem semântica com redes convolucionais profundas, convolução atrosa e crfs totalmente conectadas." Transações IEEE sobre análise de padrões e inteligência de máquina 40.4 (2017): 834-848.
3. Chen, Liang-Chieh, et al. "Repensando a convolução atrosa para a segmentação semântica de imagens." pré-impressão arXiv arXiv:1706.05587 (2017).
4. Chen, Liang-Chieh, et al. "Codificador-decodificador com convolução atrosa separável para segmentação semântica de imagem." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
5. Zhao, Hengshuang, e outros. "Rede de análise de cena de pirâmide." Anais da conferência IEEE sobre visão computacional e reconhecimento de padrões. 2017.
6. Ronneberger, Olaf, Philipp Fischer e Thomas Brox. "U-net: redes convolucionais para segmentação de imagens biomédicas." Conferência Internacional sobre Computação de Imagens Médicas e Intervenção Assistida por Computador. Springer, Cham, 2015.
7. Zhou, Zongwei, e outros. "Unet++: Uma arquitetura u-net aninhada para segmentação de imagens médicas." Aprendizado profundo em análise de imagens médicas e aprendizado multimodal para suporte à decisão clínica. Springer, Cham, 2018. 3-11.
8. Huang, Huimin, e outros. "Unet 3+: Uma rede conectada em grande escala para segmentação de imagens médicas." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.
9. Vaswani, Ashish, e outros. "Atenção é tudo que você precisa." Avanços em sistemas de processamento de informações neurais 30 (2017).
10. Wang, Zhengyang, e outros. "U-Nets não locais para segmentação de imagens biomédicas." Anais da Conferência AAAI sobre Inteligência Artificial. Vol. 34. Nº 04. 2020. See More
11. Chen, Jieneng, et al. "Transunet: Transformers são codificadores fortes para segmentação de imagens médicas." pré-impressão arXiv arXiv:2102.04306 (2021).
12. Liu, Ze, e outros. "Transformador Swin: Transformador de visão hierárquica usando janelas deslocadas." Anais da Conferência Internacional IEEE/CVF sobre Visão Computacional. 2021.
13. Cao, Hu, e outros. "Swin-unet: transformador puro semelhante ao Unet para segmentação de imagens médicas." pré-impressão arXiv arXiv:2105.05537 (2021).
14. Gao, Yunhe, Mu Zhou e Dimitris N. Metaxas. "UTNet: uma arquitetura de transformador híbrido para segmentação de imagens médicas." Conferência Internacional sobre Computação de Imagens Médicas e Intervenção Assistida por Computador. Springer, Cham, 2021. See More
15. Dosovitskiy, Alexey, et al. "Uma imagem vale 16x16 palavras: Transformers para reconhecimento de imagem em escala." arXiv preprint arXiv:2010.11929 (2020).
16. Liu, Zhuang, e outros. "Uma ConvNet para a década de 2020." pré-impressão arXiv arXiv:2201.03545 (2022).
17. Ding, Xiaohan, e outros. "Escalonando seus kernels para 31x31: revisitando o design de kernel grande em cnns." Anais da Conferência IEEE/CVF sobre Visão Computacional e Reconhecimento de Padrões. 2022.
18. Hu, Hanzhe, e outros. "Convolução de gráfico dinâmico em termos de classe para segmentação semântica." Conferência Europeia sobre Visão Computacional. Springer, Cham, 2020.
19. Meng, Yanda, e outros. "Região baseada em gráficos e agregação de limites para segmentação de imagens biomédicas." IEEE Transactions on Medical Imaging (2021).
20. Kipf, Thomas N. e Max Welling. "Classificação semi-supervisionada com redes convolucionais de grafos." pré-impressão arXiv arXiv:1609.02907 (2016).
21. Veličković, Petar, et al. "Gráfico Redes de Atenção." Conferência Internacional sobre Representações de Aprendizagem. 2018.
22. Vaswani, Ashish, e outros. "Atenção é tudo que você precisa." Avanços em sistemas de processamento de informações neurais 30 (2017).