



# Colorização de Desenhos

*L. Ferreira Paiva      H. Pedrini*

Relatório Técnico - IC-PFG-24-17  
Projeto Final de Graduação  
2024 - Dezembro

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.  
O conteúdo deste relatório é de única responsabilidade dos autores.

# Colorização de Animações

Lucca Ferreira Paiva\*

Hélio Pedrini†

## Resumo

Desenhos e animações são compostos por linhas bem definidas que separam os objetos do fundo. Dessa forma, métodos de colorização de imagens naturais não são eficientes neste cenário. Assim, várias abordagens surgiram nos últimos anos com diferentes estratégias para resolver este problema. Este trabalho investiga as principais técnicas desenvolvidas para colorização de animações, ressaltando suas semelhanças e diferenças. Além disso, ele inclui uma discussão sobre os conjuntos de dados utilizados e as metodologias de avaliação, bem como os principais desafios da área.

## 1 Introdução

Desenhos e animações são frequentemente empregados em nosso dia a dia, embora muitas vezes sem a nossa percepção. Eles estão presentes na maioria dos repositórios, em cartazes, em jogos de celulares e de computador, bem como em cinemas e salas de televisão. Eles formam um meio que possibilita liberdade para os seus artistas, permitindo que explorem novas técnicas, estilos, formas de contar histórias e de se expressarem.

As aplicações também podem ser as mais diversas, incluindo diversos setores de entretenimento, como filmes, séries de televisão e propagandas, até o uso mais variado, por exemplo, para apoio político [25]. Recentemente, as animações têm se destacado como uma ferramenta eficaz para o ensino, abrangendo desde campanhas de conscientização até ambientes mais formais. Sua capacidade de tornar o aprendizado mais dinâmico e menos monótono tem atraído atenção. Ibrahim [8] recomenda um uso mais amplo dessa tecnologia em salas de aula de escolas e universidades.

A colorização de desenhos é o processo de converter artes de linha em preto e branco em pinturas totalmente coloridas. É preciso preencher as formas definidas pelos contornos com as cores adequadas, ao mesmo tempo em que se ajusta aos níveis de iluminação, às texturas e às sombras [15]. O objetivo é criar uma imagem que tenha cores plausíveis, seja convincente e esteticamente agradável para um observador humano [29].

Este texto está organizado da seguinte forma. A Seção 2 descreve os conceitos essenciais relacionados a desenhos e animações, bem como as principais abordagens de colorizações, divididas pela forma de mídia. A Seção 3 aborda os principais conceitos de tratamento de dados e os seus conjuntos utilizados nos trabalhos da área. A Seção 4 lida com as formas de

---

\*Instituto de Computação, Universidade Estadual de Campinas, Campinas, SP, Brasil, 13083-852

†Instituto de Computação, Universidade Estadual de Campinas, Campinas, SP, Brasil, 13083-852

avaliação mais relevantes utilizadas. Por último, tem-se a Seção 5, que resume as principais conclusões do trabalho.

## 2 Principais Conceitos e Trabalhos Relacionados

Esta seção fornece uma breve descrição do conceito e do processo de desenvolvimento, além de definir a terminologia relacionada aos desenhos que será utilizada neste trabalho. Inicialmente, aborda-se a colorização de desenhos, com uma classificação dos métodos em guiados e automáticos. Em seguida, apresenta-se uma análise sobre a colorização de animações em vídeo.

Durante este trabalho, o termo *desenho* será usado de forma abrangente para se referir a imagens desenhadas manualmente no plano bidimensional (2D) para diversos fins, como mangás, histórias em quadrinhos, cartoons e outros. O termo *animação*, por sua vez, será empregado para descrever quadros desenhados manualmente e reproduzidos em sequência para criar a ilusão de movimento, frequentemente chamados de desenhos animados. Esse conceito inclui tanto animações tradicionais, como “O Rei Leão” e “Branca de Neve”, quanto animes, como “Akira”, “A Viagem de Chihiro” e “One Piece”.

A terminologia *arte de linha* ou *esboço de linha* será utilizada para designar desenhos ainda não coloridos, incluindo quadros de animações. Por outro lado, o termo *pintura* será empregado para indicar um desenho já colorido. Exemplos de artes de linha, desenhos e animações são ilustrados na Figura 1.

### 2.1 Colorização de Desenhos

Desenhos 2D destacam-se por suas características únicas, que realçam a criatividade e a personalidade das composições. Entre essas características, destaca-se o uso de expressões exageradas para efeitos artísticos, especialmente na transmissão de emoções [13]. Além disso, esse meio apresenta elevado nível de abstração e dinamismo.

As imagens são compostas por linhas explícitas, frequentemente organizadas de acordo com padrões geométricos, resultando em figuras com contornos bem definidos que servem como base para o detalhamento posterior do quadro. No entanto, a ausência de informações de textura e sombras nas artes de linha, característica dessa mídia, representa um dos principais desafios no desenvolvimento de métodos de colorização de desenhos [24].

Os métodos de colorização de desenhos são classificados em dois grupos principais: guiados e automáticos. Os métodos guiados envolvem processos que utilizam informações adicionais fornecidas pelo usuário, além da arte de linha. Essas informações podem ser fornecidas de diversas maneiras, sendo as mais comuns dicas de cor, imagens de referência ou especificações por texto.

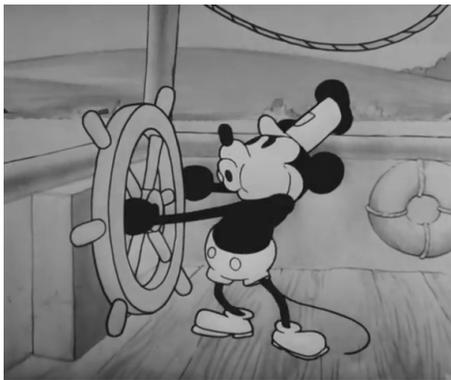
Por outro lado, os métodos automáticos operam exclusivamente com a arte de linha como entrada, sem requerer qualquer interação adicional. Um diagrama ilustrando esses métodos pode ser encontrado na Figura 2. É importante destacar que um único processo pode oferecer múltiplas modalidades de interação.



(a) Arte de linha desenhada naturalmente



(b) Desenho colorido por artista



(c) Quadro de animação



(d) Arte de linha colorida por computador com dicas de cor

Figura 1: Exemplos de (a) artes de linha, extraída de Morevna Project [17], (b) desenhos coloridos, extraído de Morevna Project [17], (c) quadro de animação de *Steamboat Willie*, extraído de Walt Disney [30] e (d) imagem colorida por método computadorizado, extraída de Morevna Project [17] e colorida por TaiZan [27].

### 2.1.1 Métodos Guiados

Métodos guiados são definidos por receberem informações adicionais do usuário para guiar o processo de colorização. Dessa forma, o artista pode ter mais controle sobre a pintura produzida, fornecendo uma paleta de cores, uma imagem de referência ou uma especificação por texto. Assim, a interação do utilizador pode reduzir significativamente a dificuldade da análise semântica e da seleção de cores adequadas. Essas razões justificam a existência de

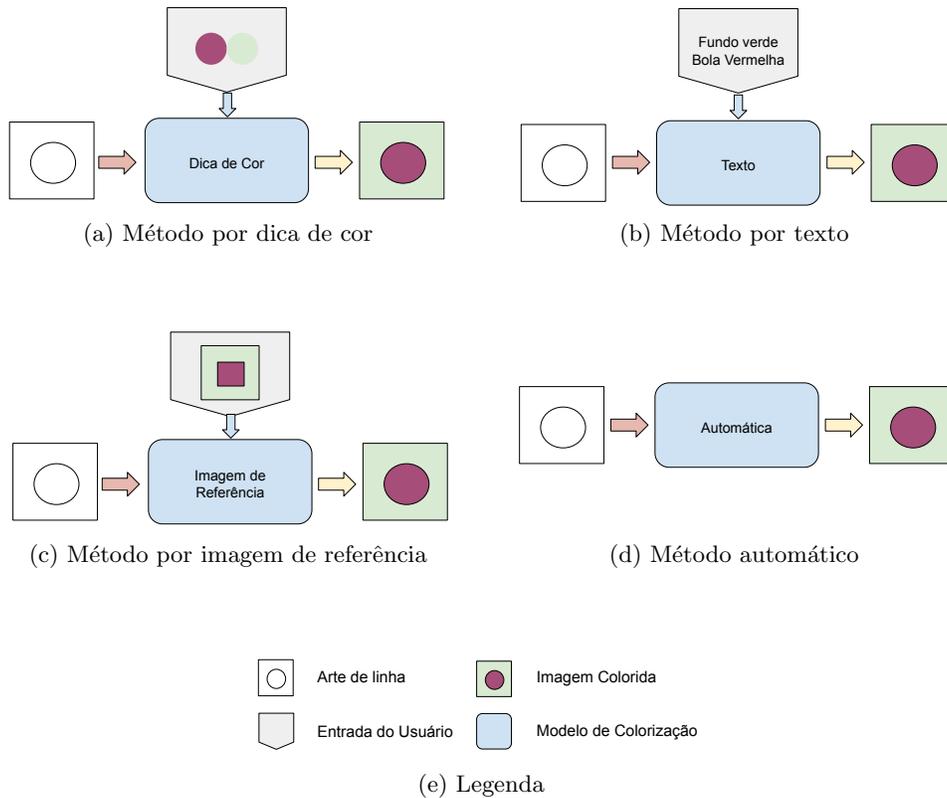


Figura 2: Tipos de colorização e entradas necessárias para cada método.

uma maior quantidade de métodos de colorização guiados do que automáticos [36]. Zhao et al. [36] dividem as abordagens guiadas em três tipos, de acordo com as estratégias de interação: orientação por imagens de referência, dicas de cor e expressões em linguagem/texto especificadas pelo usuário, destacando que um bom método semi-automático deve ser interativo e revisável.

*Dicas de cor* permitem que o usuário especifique as cores desejadas na pintura final e o local em que devem ser aplicadas. Assim, estes métodos possibilitam maior participação no processo de colorização, permitindo que o artista tenha mais liberdade e consiga expressar sua criatividade. Zhao et al. [36] definiram funcionalidades que uma rede baseada em dicas de cor deve ser capaz de cumprir: determinar automaticamente o contorno de áreas fechadas, transicionar automaticamente entre cores e aplicar detalhes, sombra e textura.

Para a colorização de imagens com o uso de cores fornecidas pelo usuário, Zhang et al. [33] e Lee e Lee [12] propuseram uma abordagem em duas etapas: esboço e refinamento. A partir de uma arte de linha e algumas indicações de cor, a primeira fase é treinada para determinar a composição geral e selecionar cores mais vivas. No trabalho de Zhang et al. [33], essa etapa inicial visa seguir estritamente as entradas fornecidas, colorindo de maneira “agressiva” para maximizar a diversidade de cores.

Na etapa de refinamento, áreas incorretamente coloridas são identificadas e corrigidas. O treinamento dessa fase foca nos detalhes, utilizando dados que incluem exemplos com vazamentos de cor, imagens borradas e colagens de partes de outras figuras. O método permite intervenções em três momentos do processo: na imagem de entrada, no rascunho e na saída final, oferecendo aos artistas a possibilidade de revisar e ajustar artefatos gerados. A Figura 3 ilustra os resultados dessa rede, comparando-os com outros modelos, como o TaiZan [27], um site online de colorização.

Por sua vez, Lee e Lee [12] enfatizam a importância dos dados de treinamento, destacando a distribuição de linhas como um fator crítico para a eficiência das redes. Além disso, o método deles é projetado para criar pinturas de forma totalmente automática, sem depender de imagens de referência.

Na abordagem baseada em *imagens de referência*, é necessário fornecer uma imagem semelhante à que será colorida, permitindo que o modelo estabeleça conexões entre ambas e aplique as cores com base na figura de referência. Para que isso funcione adequadamente, é essencial que as ilustrações apresentem similaridade semântica, ou seja, compartilhem elementos como objetos, personagens ou cenários.

Essa técnica está intimamente ligada ao conceito de transferência de estilo, em que um desenho colorido base é utilizado em conjunto com uma arte de linha, e o modelo transfere o estilo de pintura e, possivelmente, as cores de uma imagem para outra. Embora a necessidade de uma imagem de referência colorida possa ser uma desvantagem, especialmente em casos onde sua obtenção é desafiadora [33], esse método apresenta a vantagem de oferecer maior consistência de cor ao lidar com múltiplas imagens.

Alguns dos pioneiros na área de colorização foram Sýkora et al. [26] em 2004, que utilizaram métodos de aproximação antes da era do aprendizado profundo [24]. Os autores empregaram a técnica de amostragem baseada em fragmentos para extrair regiões de interesse e, por meio de mapeamento reverso, fizeram a correspondência das características entre os quadros. Esse processo permitiu o pareamento das características e a aplicação de cores com base na semelhança entre a arte de linha e o quadro de referência. Em seguida, Qu et al. [19] se basearam nesse trabalho, mas focaram em imagens de mangá. Eles fizeram uso de texturas e intensidades para facilitar a colorização das figuras.

Mais recentemente, métodos baseados em aprendizado de máquina começaram a ser amplamente utilizados, como o trabalho de Hensman e Aizawa [7], que adota uma abordagem em que apenas uma imagem é usada para o treinamento, podendo ser considerada uma imagem de referência. O modelo emprega a cGAN (rede adversária generativa condicional), desenvolvida por Isola et al. [10], para realizar a primeira colorização, utilizando uma rede destinada à colorização de imagens naturais.

A imagem alvo é então segmentada e colorida em quatro fases. Na primeira fase, é selecionada uma cor para cada segmento. A segunda fase pode aumentar a saturação para gerar resultados mais agradáveis visualmente. A terceira etapa consiste na quantização das cores, com o objetivo de reduzir ruídos e padronizar as cores. Por fim, são aplicados efeitos de sombreamento. Uma desvantagem desse método é que ele exige treinamento adicional sempre que novos detalhes, como personagens ou objetos, são inseridos nas imagens.

Furusawa et al. [6] aplicam a estratégia de referência para colorir páginas de mangá, onde o usuário fornece sugestões de cores e a rede aplica essas cores nos painéis da página.

O trabalho se baseia na rede desenvolvida por Iizuka et al. [9], cujo modelo consiste em quatro redes: uma para características globais, uma para características de baixo nível, uma para colorização de nível médio e uma para colorização final. Para adaptar a estrutura a esse novo contexto, os autores adicionaram três componentes voltados para o escopo de desenhos: primeiro, um alavancador de características de cores da imagem de referência; segundo, realizaram treinamento do ramo de classificação utilizando nomes de personagens; e por último, fizeram um treinamento adicional com perda adversária para uma colorização mais vívida.

Na Figura 3 está uma colorização feita pela rede, mas no caso mencionado, o método acaba falhando em produzir imagens com alta qualidade de cor e vivacidade, resultando em uma colorização predominantemente cinza. Vale ressaltar que essa imagem não foi gerada nas condições originais do trabalho [33].



Figura 3: Imagem extraída de [33], para a visualização de resultados de diversas redes para (a) um personagem simples, (b) um homem com um animal detalhado, (c) ampliação do animal, (d) uma garota com um vestido de padrão complexo e (e) ampliação da decoração no cabelo. Note que *Comicolorization* [6] não usa entrada de usuário. Primeira linha: *xueshiwang*, segunda linha: *style2paints* e quarta linha: *KimonoAoi*.

Dai et al. [3] utilizam uma arquitetura baseada em U-Net com características semânticas adicionais extraídas pelo CLIP (Pré-treinamento Contrastivo de Linguagem-Texto) [20]. O

modelo é especializado em colorir imagens que apresentam grandes deformações ou que estão cobertas ou segmentadas por outras linhas. O processo começa com o cálculo do fluxo óptico entre as artes de linha, para obter uma colorização preliminar, seguida pela aplicação das cores de uma imagem de referência. Em seguida, o módulo de pareamento de inclusão refina o resultado para segmentos mais detalhados. Esta abordagem foi inspirada no método tradicional utilizado por artistas digitais, que começam colorindo o fundo e as partes mais simples, para depois focar nas regiões mais delicadas. O resultado dessa técnica pode ser visualizado na Figura 4.

Já Zhang et al. [32] optam por uma rede U-Net residual para a transferência de estilo, em conjunto com uma AC-GAN (rede adversária generativa com classificador auxiliar). Dois decodificadores guiados são introduzidos para evitar a atenuação de cores e garantir melhor qualidade na colorização. No entanto, devido ao uso de um modelo VGG16/19, essa rede enfrenta algumas limitações, como o tamanho excessivo do modelo, dificuldades no controle das cores e a restrição do tamanho da imagem final, que é limitada a  $256 \times 256$  pixels [12].

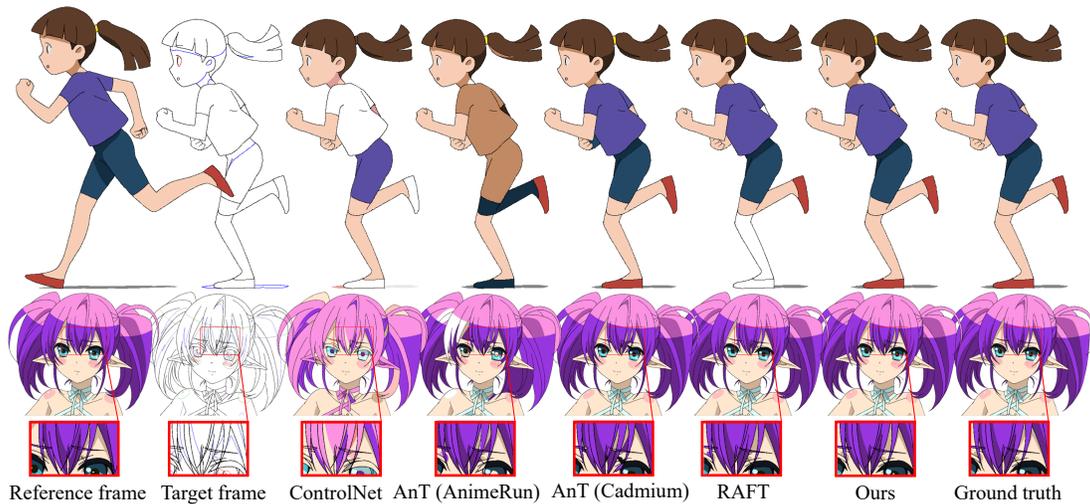


Figura 4: Imagem extraída de Dai et al. [3] que mostra a eficácia do método para efeitos de oclusão e movimento de imagens simples.

Métodos de *colorização baseada em texto* oferecem a vantagem de serem mais acessíveis, proporcionando uma interface mais simples para usuários não profissionais. No entanto, essa abordagem apresenta a desvantagem de ser desafiadora em termos de fazer o mapeamento entre as instruções linguísticas e a arte de linha.

Dois dos principais trabalhos na área foram propostos por Zou et al. [39] e Zou et al. [37], ambos destacando a interação baseada em linguagem para colorização e a aplicação de compreensão semântica na identificação e coloração de objetos. Ambos os textos começam segmentando as linhas e mapeando as instâncias para utilizar redes cGAN, colorindo cada instância ou região de fundo. O primeiro trabalho sugere a combinação com métodos tradicionais de colorização para criar um sistema multimodal, enquanto o segundo descreve especificamente a arquitetura do LUCSS, composta por três módulos (segmentação, legenda

e colorização interativa), detalhando como eles colaboram no processo de colorização com base em edição textual. Outro trabalho relevante nessa área é Kim et al. [11], cuja rede é treinada para receber rótulos. Essa abordagem é ilustrada na Figura 5.

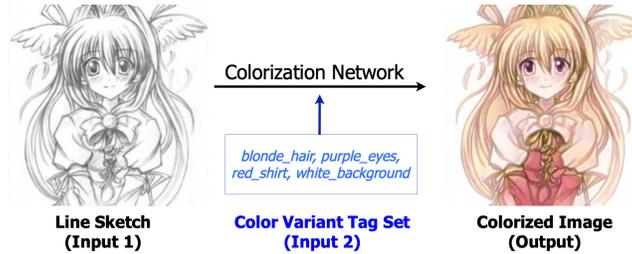


Figura 5: Imagem extraída de Kim et al. [11] que mostra uma forma geral das redes baseadas em texto.

### 2.1.2 Métodos Automáticos

A colorização automática é definida como o processo que recebe apenas a arte de linha, de forma que os métodos consigam perceber suas características e concatenar texturas, sombras e, principalmente, cores, de forma imperceptível, criando uma imagem colorida esteticamente agradável [12]. Existem duas razões específicas que tornam esse processo difícil, conforme apontado por Zhao et al. [36]. A primeira é que uma colorização meticulosa exige uma identificação semântica precisa de todos os alvos e componentes presentes na imagem. A segunda é que as abordagens precisam produzir resultados visualmente interessantes e consistentes, ao mesmo tempo em que mantêm um detalhamento suficiente.

Liu et al. [15] foram alguns dos primeiros a trabalhar com a colorização automática, desenvolvendo uma SGRU (*U-Net Residual com Gate de Swish*) como um modelo totalmente autônomo, dispensando a necessidade de intervenção manual. Eles utilizaram uma arquitetura de rede U-Net, onde o caminho de codificação captura as características e progressivamente reduz a resolução da figura de entrada, enquanto o caminho de decodificação aumenta novamente. A inclusão de blocos residuais melhora a aprendizagem de características, permitindo que o sistema retenha informações e faça ajustes finos, especialmente úteis em imagens com detalhes, como esboços de anime, que eram o foco do projeto.

O Auto-painter [16] é um modelo de aprendizado supervisionado que, dado um esboço em preto e branco, pode gerar uma imagem colorida com base em pares de esboço e imagem colorida fornecidos ou apenas com os dados de treinamento. Utilizando uma rede neural profunda com *feed-forward* como gerador, o modelo visa uma resposta rápida. O esboço é recebido como entrada, e a rede gera uma imagem colorida no mesmo nível de resolução da arte de linha, pixel a pixel. A rede usa uma cGAN, onde tanto o gerador quanto o discriminador são condicionados pela imagem de entrada, visando melhorar o desempenho supervisionado. Com o auxílio de uma rede U-Net, que contém apenas a unidade de codificação, a imagem gerada é fornecida ao discriminador, que tem a tarefa de classificar se o par arte de linha e imagem são reais ou falsos. Alguns dos resultados podem ser vistos na Figura 6.

Em uma abordagem diferente, Aizawa et al. [1] optam por focar em uma parte específica do problema de colorização, utilizando uma rede U-net para detectar as regiões de esclera, a parte branca dos olhos. Eles combinam essa rede com o *PaintsChainer* [21] para reforçar o processo de colorização e aumentam a precisão do modelo por meio de segmentação semântica [36]. Ao concentrar-se nessa região específica, conseguem gerar resultados eficazes, demonstrando que a abordagem pode ser vantajosa quando aplicada a partes bem definidas de uma imagem.

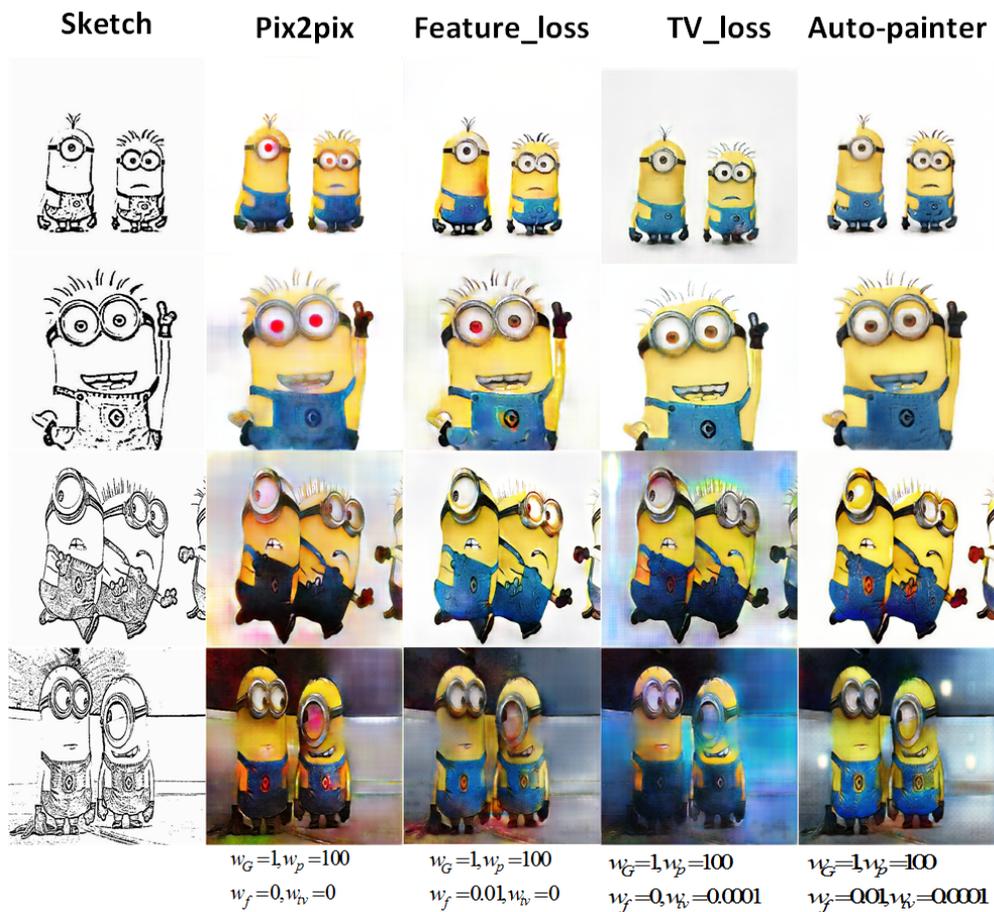


Figura 6: Imagem extraída de Liu et al. [16] que mostra a colorização automática do modelo *Auto-painter* e *Pix2Pix* [16, 10] com diversas configurações.

## 2.2 Colorização de Animações em Vídeo

Animações são construídas a partir de várias imagens exibidas sucessivamente, criando uma ilusão de movimento quando reproduzidas a uma velocidade específica [8]. Esse processo, embora dê a sensação de continuidade, é bastante trabalhoso e exige um tempo considerável para ser completado. A criação de uma animação envolve diversas etapas e frequentemente

requer a colaboração de múltiplas equipes especializadas. Entre essas etapas, destacam-se o desenho dos quadros-chave, que são as poses principais e que geralmente são realizados pelos artistas mais experientes e renomados. Em seguida, criam-se as imagens intermediárias, que preenchem os intervalos entre os quadros-chave. Posteriormente, outras fases como colorização, criação do fundo, adição de efeitos visuais e outras, completam o processo de produção da animação.

A colorização das linhas de contorno é frequentemente considerada uma tarefa entediante, repetitiva e de baixa compensação [28], além de ser um processo dispendioso. Por conta disso, tornou-se crucial o desenvolvimento de ferramentas que possam dinamizar, acelerar e reduzir o custo dessa etapa. Nos últimos anos, abordagens baseadas em aprendizado de máquina e aprendizado profundo têm ganhado cada vez mais destaque na área. Apesar de a colorização ser um problema antigo, a aplicação dessas técnicas na colorização de animações em vídeo ainda não foi explorada de forma tão abrangente quanto na colorização de imagens naturais, nem com a mesma profundidade que a colorização de desenhos [36]. No entanto, ambas as áreas compartilham diversas similaridades, o que torna o estudo da colorização de animações um campo promissor para a aplicação dessas tecnologias.

Uma das maiores dificuldades ao lidar com animações é a falta de continuidade entre quadros consecutivos, o que ocorre devido à tendência de se desenharem movimentos grandes e extravagantes. Dessa forma, os corpos dos personagens “saltam” de uma posição para outra, ao invés de se moverem de maneira suave. Isso significa que as imagens não mantêm uma continuidade em nível de pixel entre os quadros [23], o que dificulta a colorização e o processo de renderização suave ao longo do tempo. Além disso, outro aspecto relevante é que, em muitas animações, os personagens mudam de posição a cada dois quadros, um conceito conhecido como “animar em dois”, o que pode afetar negativamente o treinamento e a entrada das redes de processamento, tornando o modelo mais complexo e exigente em termos de dados.

A coerência temporal é um dos maiores desafios na colorização de vídeos, pois exige que o mesmo objeto mantenha suas cores consistentes ao longo de vários quadros. Isso implica que não é possível simplesmente aplicar métodos de colorização de desenhos sem ajustes, pois o uso desses métodos resultaria em quadros vizinhos inconsistentes, comprometendo a fluidez e a continuidade visual da animação. Além disso, as movimentações bruscas dos personagens e objetos dificultam a aplicação de técnicas como os algoritmos de campo óptico [26, 23], amplamente utilizados na colorização de vídeos naturais, para detectar e rastrear os movimentos entre os quadros. Essa limitação se torna ainda mais pronunciada em animações, onde o estilo artístico e a falta de transições suaves entre quadros tornam essas abordagens menos eficazes [26, 23].

Thasarathan et al. [28] foram um dos pioneiros no estudo de métodos voltados para a colorização de animações [34], apresentando o método de Colorização de Vídeo Automática com Coerência Temporal (ATCVC), baseado em uma GAN inspirada no modelo de Isola et al. [10]. A principal estratégia adotada nesse método é utilizar o quadro previamente colorido como uma condição para o desenho atual, com o objetivo de aumentar a consistência entre os quadros. Além disso, tanto o quadro anterior quanto o atual são alimentados no discriminador, com o intuito de garantir a coerência temporal. Para a colorização do primeiro quadro, em que não há um quadro anterior para servir de base, o modelo utiliza

uma imagem em branco [28].

Contudo, há três críticas principais ao método [34, 22, 23]. A primeira refere-se ao desalinhamento semântico entre quadros: como a distribuição espacial dos objetos pode variar, isso pode causar uma desorientação na colorização, sendo mais visível quando ocorrem mudanças significativas. A segunda crítica destaca o fato de que as informações são recalculadas em cada nova instância, o que faz com que a correlação entre o quadro anterior e o sucessor seja ignorada, mesmo que estejam altamente relacionadas. A última crítica envolve a propagação de erros, ou seja, erros introduzidos na colorização de um quadro podem afetar os quadros subsequentes, gerando uma propagação de falhas na colorização [23].

Apesar dessas limitações, o trabalho de Thasarathan et al. [28] foi um marco importante na área, incentivando diversas melhorias subsequentes e servindo de base para o desenvolvimento de métodos de ponta, como os apresentados por Shi et al. [23], Shi et al. [22] e Zhang et al. [34].

O artigo de Zhang et al. [34] apresenta o método *Line Art Correlation Matching Feature Transfer Network* (LCMFTN) para a colorização automática de animações em arte de linha. Este método utiliza um modelo de transferência de características baseado em correspondência de correlação (CMFT), que se aproveita das informações temporais para reduzir cintilações ou mudanças bruscas de cor, garantindo, assim, a consistência temporal das cores entre quadros consecutivos. O uso do CMFT permite a identificação de diferentes tipos de características semânticas (como estrutura e formas) e texturais (detalhes finos), utilizando o pareamento dessas características para identificar áreas similares entre os quadros e transferir as cores de maneira mais fiel possível.

Esse processo ajuda a garantir que as áreas da animação que compartilham atributos visuais recebam a mesma paleta de cores ao longo dos quadros, promovendo maior estabilidade visual e naturalidade, ao mesmo tempo que preserva o estilo original da linha. As cores são aplicadas de maneira coesa, o que também evita a reprodução de erros comuns em colorização de animações, como desalinhamento de cores e mudanças indesejadas entre os quadros. Os resultados do método LCMFTN podem ser visualizados na Figura 7, que compara essa arquitetura com as abordagens de Thasarathan et al. [28], Isola et al. [10] e Liao et al. [14].

O modelo criado por Shi et al. [23] se baseia em uma estrutura GAN de aprendizado não supervisionado para colorir vídeos de arte de linha, utilizando uma combinação de características globais e locais extraídas de imagens de referência. A rede consiste em duas sub-redes: uma de transformação de cor e uma de refinamento. Este modelo é embasado no trabalho similar de Shi et al. [22], que utiliza uma arquitetura com atributos semelhantes, mas voltada para um uso eficiente de poucas referências.

A primeira rede, de transformação de cor, usa um gerador que aplica cores com base no estilo global e nos detalhes locais da cor, juntamente com um discriminador que ajuda a avaliar e aprimorar a qualidade da colorização. O gerador recebe uma imagem de arte de linha e um mapa de campo de distância alvo de duas imagens de referência como entrada, produzindo a imagem colorida alvo. O mapa, que reflete a relação entre a imagem de entrada e a colorida de referência, auxilia o gerador a identificar áreas de mesma cor local com base na proximidade e relevância entre as figuras fornecidas.

A rede de refinamento ajusta o resultado preliminar gerado pela etapa de transformação de cor, corrigindo possíveis inconsistências e aprimorando detalhes de cor. Esta fase trabalha para melhorar a continuidade de cores entre os quadros consecutivos, garantindo uma transição mais suave e minimizando falhas na aplicação da cor. Assim, a rede visa aplicar cores com consistência temporal e precisão, mesmo diante de mudanças visuais entre os quadros.

Os métodos de colorização de vídeos de animação se destacam por seu foco na consistência temporal, que é a principal dificuldade em relação à colorização de desenhos. A utilização de imagens base como referência tem mostrado melhorias nos resultados dos métodos de estado da arte. Sistemas de atenção e a captura de características globais das referências parecem aprimorar a aplicação de cores, especialmente em detalhes específicos. No entanto, esta continua sendo uma área muito recente, repleta de desafios, como a introdução de personagens novos, para os quais não há referências disponíveis [28]. Ainda não há um consenso claro sobre a melhor forma de implementação, com técnicas como redes generativas adversariais (GANs), mecanismos de atenção e o uso de múltiplas imagens de referência ainda sendo amplamente exploradas [24].

### 3 Conjuntos de Dados

É difícil trabalhar com desenhos porque eles precisam ser desenhados e coloridos manualmente, ao passo que obter imagens do mundo real é mais simples, pois podem ser facilmente capturadas com uma câmera e convertidas para preto e branco. Dessa forma, dados de desenhos não são tão acessíveis quanto imagens do mundo real. Além disso, o ideal seria ter acesso aos pares de arte de linha originais e suas respectivas pinturas coloridas, mas adquirir a imagem original é um grande desafio. A alternativa passa a ser extrair as artes de linha a partir de quadros já coloridos. Isso revela uma das maiores dificuldades da área: a falta de dados, especialmente a ausência de conjuntos de dados consolidados [36]. Como resultado, muitos estudos acabam criando suas próprias bases de dados de acordo com o foco específico de cada pesquisa. Embora isso permita a criação de conjuntos de dados mais especializados, também torna a comparação entre técnicas mais complexa. Ademais, o grande número de estilos de animações existentes dificulta a criação de um conjunto de dados diverso, que seja composto por imagens de alta qualidade com os pares de imagens em preto e branco e coloridas.

Uma solução adotada por vários autores é extrair as artes de linha de imagens já coloridas. Existem múltiplos métodos desenvolvidos com essa finalidade, sendo os mais comuns o *XDoG* [31] e o *PaintsChainer* [27], que foram utilizados, entre outros, por Zhang et al. [33], Liu et al. [16] e Ci et al. [2]. Neste contexto, é importante destacar o trabalho de Lee e Lee [12], que aborda o problema de a performance de colorização ser melhor em artes de linha geradas artificialmente do que nas desenhadas por artistas. Os autores argumentam que isso ocorre devido a um sobreajuste da entrada das artes de linha, causado pela diferença na disposição dos pixels entre as imagens desenhadas e as extraídas por algoritmos. Esse fenômeno pode ser observado na Figura 8, que mostra como as imagens feitas por humanos não possuem textura no fundo, ao contrário das extraídas pelos métodos mencionados.

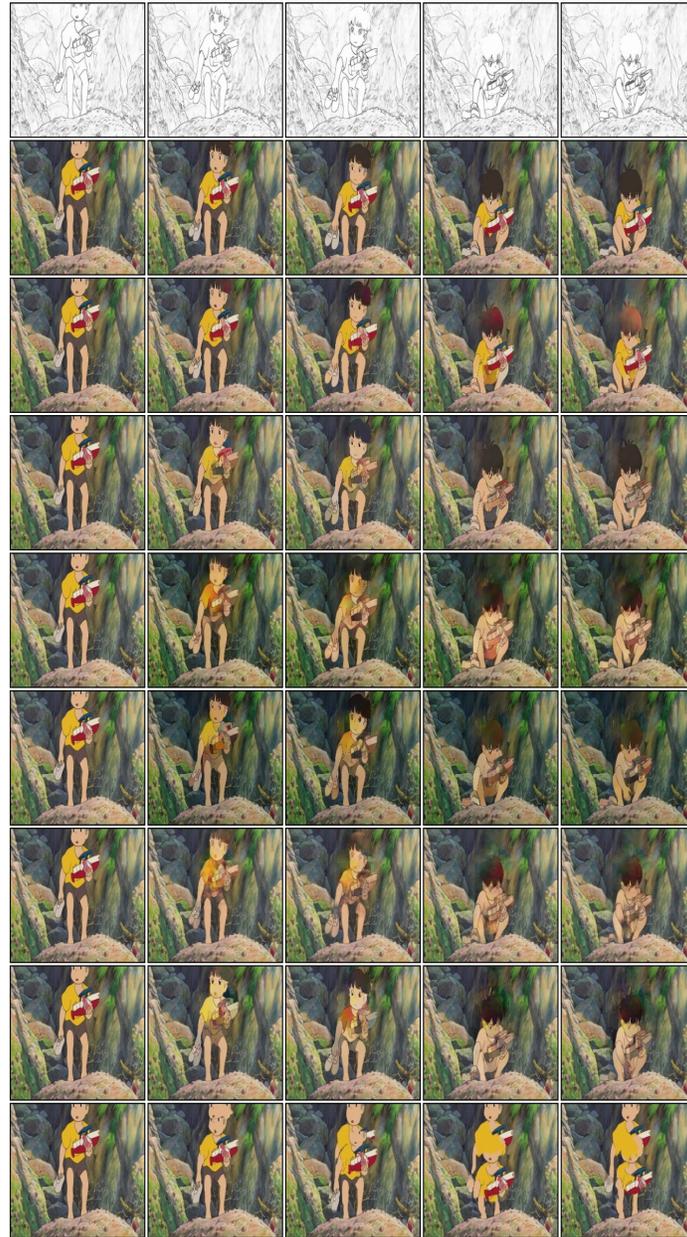


Figura 7: Imagem extraída de Zhang et al. [34] que mostra os resultados de diferentes colorizações. Do topo para baixo tem-se: a arte de linha obtida, a imagem original, resultado de LCMFTN [34], LCMFTN (sem CMFT), TCVC (com a perda do trabalho) [28], TCVC, Pix2Pix [10] (com a referência e perda do trabalho), *Pix2Pix* (com referência), *DeepAnalogy* [14]. A primeira coluna representa a imagem colorida de referência, e as sucessivas são as coloridas baseadas na imagem referência. A sequência é do filme *Ponyo*.

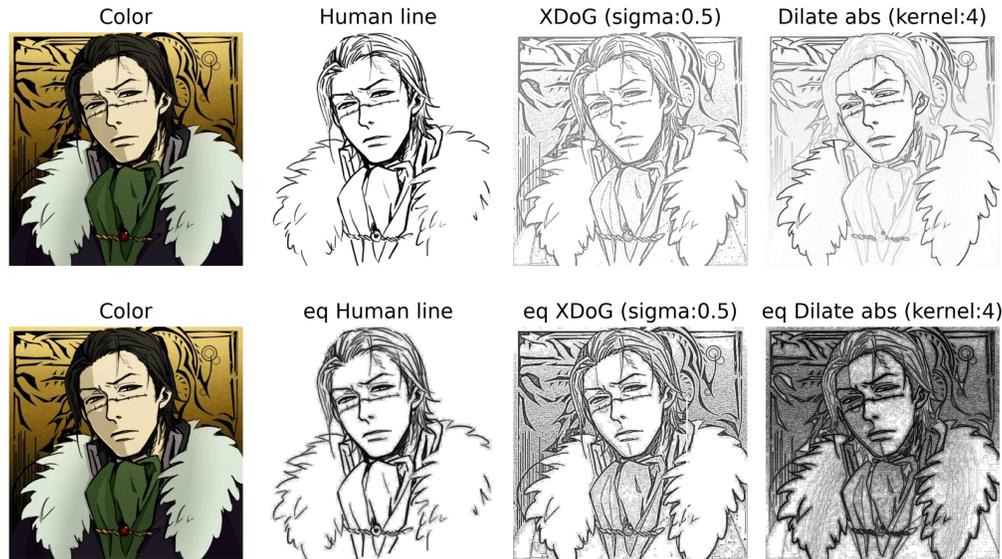


Figura 8: Imagem extraída de Lee e Lee [12] que mostra os diferentes resultados para diferentes parâmetros para a extração de linhas, utilizando o *XDoG* e *Dilate abs*, este um criado no trabalho.

Uma das maiores fontes de imagens para colorização são os sites e plataformas online dedicadas ao compartilhamento de desenhos. Um dos maiores exemplos é o *Danbooru* [4], em suas diversas edições, que é um repositório online de imagens, especialmente focado em artes e ilustrações de estilo anime e mangá. No site, usuários podem compartilhar e visualizar figuras, sendo notável pela riqueza dos dados, pois as imagens são rotuladas com detalhes como personagens, temas, cores e estilos. Este conjunto de dados foi utilizado por Liu et al. [15] e Zhang et al. [33]. Contudo, também existem críticas à plataforma, como o fato de suas imagens conterem muito ruído e traços auxiliares usados no processo de desenho, que não fazem parte da arte final [12, 2], o que exige um cuidado especial na filtragem das imagens selecionadas. Outras alternativas podem ser consideradas, como *niconico-seiga* [18] e *E-Shuushuu* [5], que foram utilizadas por Furusawa et al. [6] e Lee e Lee [12]. Além disso, outros conjuntos de dados como *SketchyScene* [38] também foram utilizados por Zou et al. [37] e Zou et al. [39].

No entanto, na maioria dos casos, ainda é necessário criar ou ao menos adaptar outro conjunto de dados para atender às necessidades específicas de cada trabalho. Por exemplo, Hensman e Aizawa [7] utilizam duas fontes específicas: *Monster Beat*, um mangá e *Morevna*, um desenho baseado no mesmo estilo. Outro aspecto importante é a própria qualidade da cor das imagens, sendo muitas vezes ideal remover aquelas com baixa saturação para melhorar os resultados da colorização [6]. Esse processo de curadoria de dados, que inclui a seleção de imagens com cores vibrantes e a remoção de ruídos ou imagens de baixa qualidade, é fundamental para o sucesso dos modelos de colorização, especialmente em contextos de animações e ilustrações artísticas.

No campo das animações, os conjuntos de dados são ainda menos padronizados, o que resulta em uma grande variedade de fontes utilizadas em cada estudo. Comumente, as fontes incluem animes populares, como *Dragon Ball*, que foi influenciado pelo trabalho de Thasarathan et al. [28]. Outra fonte frequente são os filmes do renomado diretor japonês Hayao Miyazaki, amplamente utilizado em estudos como os de Zhang et al. [34], Shi et al. [22] e Shi et al. [23].

## 4 Formas de Avaliação

Avaliar a qualidade de imagens geradas continua sendo um desafio, especialmente no contexto de colorização de desenhos [35]. Diversas metodologias foram propostas para medir os erros mais comuns e evitar sua ocorrência. Essas abordagens podem ser classificadas em subjetivas ou objetivas, dependendo de como a qualidade é medida. A seguir, listam-se alguns dos erros mais comuns em colorização de imagens:

- Erros de cor: Ocorrem quando um objeto é colorido com a cor errada, como um sol azul ou rostos humanos verdes, prejudicando a fidelidade e a naturalidade da imagem.
- Sangramento de cor: Refere-se ao fenômeno quando as cores vazam para áreas adjacentes. Por exemplo, a cor da pele pode vazar para o fundo, criando uma transição indesejada entre áreas distintas da imagem.
- Borrados e distorções: Aparecem quando a imagem resultante tem regiões borradas ou com muita textura, como em uma pintura com aquarela, ou quando as linhas estruturais estão comprometidas devido à distorção excessiva das áreas coloridas.
- Falta de cor e saturação: Ocorre quando a imagem colorida apresenta cores com pouca intensidade ou saturação, geralmente com tons apagados ou dominados por tons de cinza, o que resulta em uma estética sem vida ou muito aquosa.

Para ilustrar os resultados de alguns métodos, pode-se observar a Figura 9, extraída de [36], que apresenta os resultados de *PaintsChainer* [27], *Tag2Pix* [11] e *style2paintsV3* [33], utilizando o conjunto de dados Danbooru 2017 [4]. É importante notar que, ao se utilizar a mesma imagem de arte de linha como entrada, diferentes métodos podem gerar cores distintas para as mesmas áreas. Além disso, essas técnicas também podem produzir imagens com variações significativas em termos de conteúdo e qualidade, como texturas, brilho e sombras.

Outro ponto relevante é que a interação com cada um desses métodos é única, de modo que a imagem apresentada ilustra apenas as principais características de cada técnica, sem oferecer uma comparação direta entre elas. Isso destaca um dos problemas centrais na área: muitos métodos são desenvolvidos para conjuntos de dados, estilos ou aplicações específicas, o que torna difícil compará-los de forma justa em um único conjunto de teste [36].

As principais métricas quantitativas utilizadas para medir a qualidade das imagens geradas incluem a Distância de Interceptação de Fréchet (FID), Similaridade Estrutural (SSIM), Relação Sinal-Ruído de Pico (PSNR) e Erro Médio Quadrático (MSE). A métrica FID mede

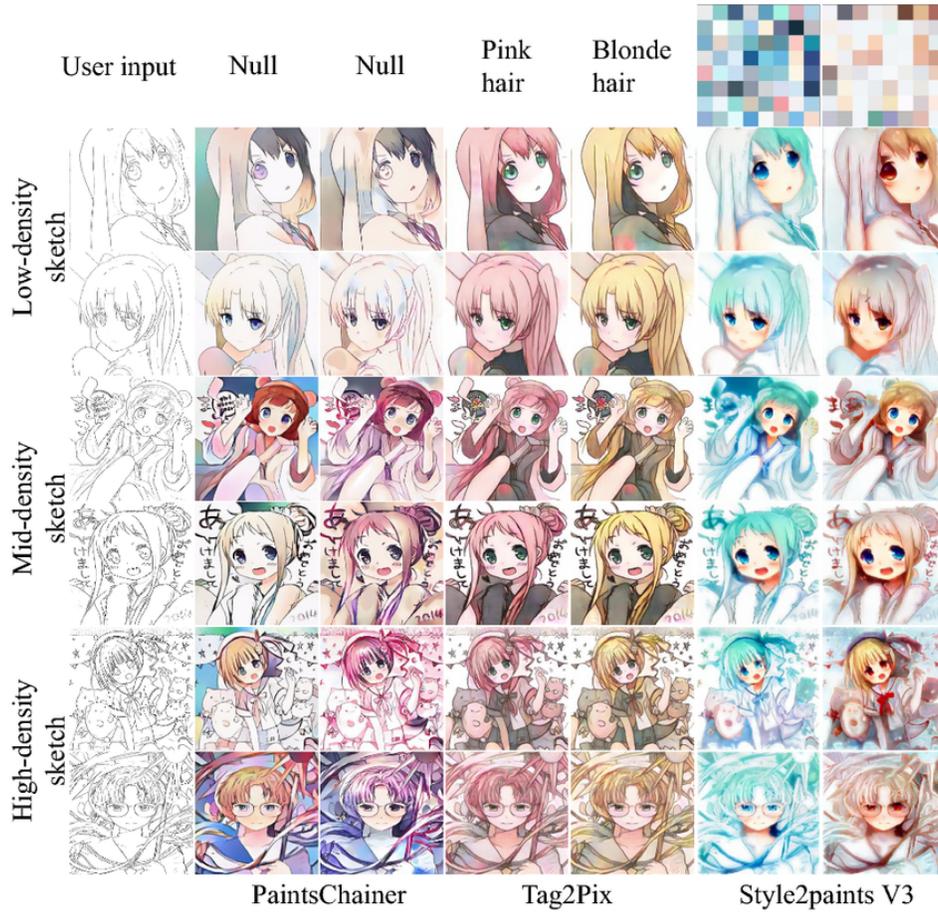


Figura 9: Imagem extraída de Zhao et al. [36], exibindo diversos métodos de colorização de desenhos com diferentes imagens.

a similaridade entre dois conjuntos de dados e, geralmente, está correlacionada com o julgamento humano sobre a qualidade visual de uma imagem. No entanto, nenhuma dessas métricas é especificamente projetada para o uso em modelos de colorização, pois elas não consideram a cor e as informações presentes nas imagens fornecidas como entrada [12, 15].

Para resolver essa limitação, Lee e Lee [12] propõem a C-FID (Colorization FID), uma adaptação da FID desenvolvida especificamente para avaliar colorizações. A C-FID leva em consideração a dica de cor (a informação de cor fornecida como entrada ao modelo), permitindo uma comparação mais precisa da qualidade das imagens coloridas, considerando a fidelidade das cores ao que foi inicialmente fornecido na arte de linha.

As métricas qualitativas, que geralmente envolvem avaliações realizadas por humanos, também são amplamente utilizadas, especialmente a Pontuação Média de Opinião (MOS), que visa tornar o processo de avaliação mais objetivo. Por exemplo, Lee e Lee [12] conduziram um estudo com 118 participantes, pedindo-lhes para avaliar várias imagens geradas por diferentes métodos em uma escala de 1 (ruim) a 5 (excelente), com base na qualidade

percebida. No entanto, avaliações que envolvem julgamentos humanos sobre imagens coloridas são lentas e caras, o que dificulta sua aplicação em larga escala e limita a eficiência desses métodos de avaliação [29].

## 5 Conclusões

A área de colorização de desenhos e animações tem avançado de maneira contínua, com melhorias constantes nos métodos utilizados. Neste artigo, diversos métodos foram analisados e resumidos, levando em consideração suas especialidades, teorias, modelos de arquitetura, aplicações e conjuntos de dados. O objetivo deste trabalho é contribuir para o entendimento das técnicas que têm sido relevantes até o momento e continuam a ser exploradas. A ausência de uma abordagem definitiva aponta para as amplas oportunidades de pesquisa que ainda existem nesse campo. Uma das principais áreas de avanço poderia ser o desenvolvimento de um conjunto de dados abrangente, que inclua desenhos e suas versões coloridas, além da padronização e criação de métodos de avaliação consolidados, que possam ser utilizados em futuras redes e técnicas de colorização.

Nossa capacidade de visualizar cores em imagens em preto e branco é natural, o que torna o processo de colorização um desafio que exige grande talento artístico. Para colorir e sombrear um desenho de forma convincente, esteticamente agradável e que transmita sua essência, é necessário não apenas habilidade técnica, mas também uma sensibilidade criativa refinada. A evolução dos métodos de colorização visa facilitar esse processo, tornando-o mais acessível e permitindo que mais pessoas se tornem artistas. A tecnologia, nesse contexto, não apenas simplifica a criação de desenhos, mas também estimula a inovação e a expressão artística, ampliando as possibilidades criativas e democratizando o acesso à arte.

## Referências

- [1] M. Aizawa, Y. Sei, Y. Tahara, R. Orihara, and A. Ohsuga. Do You Like Sclera? Sclera-region Detection and Colorization for Anime Character Line Drawings. *International Journal of Networked and Distributed Computing*, 7, 01 2019. 9
- [2] Y. Ci, X. Ma, Z. Wang, H. Li, and Z. Luo. User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks. In *26th ACM International Conference on Multimedia*, page 1536–1544, New York, NY, USA, 2018. Association for Computing Machinery. 12, 14
- [3] Y. Dai, S. Zhou, Q. Li, C. Li, and C. C. Loy. Learning Inclusion Matching for Animation Paint Bucket Colorization, 2024. URL <https://arxiv.org/abs/2403.18342>. 6, 7
- [4] Danbooru Community. Danbooru2021: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset. <https://www.gwern.net/Danbooru2021>, 2021. Acessado em: 8 de novembro de 2024. 14, 15
- [5] E-Shuushuu Community. E-Shuushuu—Kawaii Image Board. <https://e-shuushuu.net/>, 2018. Acessado em: 8 de novembro de 2024. 14

- [6] C. Furusawa, K. Hiroshiba, K. Ogaki, and Y. Odagiri. Comicolorization : Semi-automatic Manga Colorization. *CoRR*, abs/1706.06759, 2017. 5, 6, 14
- [7] P. Hensman and K. Aizawa. cGAN-based Manga Colorization Using a Single Training Image. *CoRR*, abs/1706.06918, 2017. URL <http://arxiv.org/abs/1706.06918>. 5, 14
- [8] B. M. A. Ibrahim. The Effect of Animation on The Society During The COVID-19 Pandemic A literature Review. *Journal of Arts and Humanities*, 10:63–71, Sept. 2023. 1, 9
- [9] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Let there be Color! Joint End-to-End Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. *ACM Transactions on Graphics*, 35(4), July 2016. 6
- [10] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *CoRR*, abs/1611.07004, 2016. URL <http://arxiv.org/abs/1611.07004>. 5, 9, 10, 11, 13
- [11] H. Kim, H. Y. Jhoo, E. Park, and S. Yoo. Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss. *CoRR*, abs/1908.05840, 2019. URL <http://arxiv.org/abs/1908.05840>. 8, 15
- [12] Y. Lee and S. Lee. Automatic Colorization of Anime Style Illustrations Using a Two-Stage Generator. *Applied Sciences*, 10(23), 2020. 4, 5, 7, 8, 12, 14, 16
- [13] S. Li, S. Zhao, W. Yu, W. Sun, D. N. Metaxas, C. C. Loy, and Z. Liu. Deep Animation Video Interpolation in the Wild. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6583–6591, 2021. 2
- [14] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang. Visual Attribute Transfer through Deep Image Analogy. *CoRR*, abs/1705.01088, 2017. 11, 13
- [15] G. Liu, X. Chen, and Y. Hu. Anime Sketch Coloring with Swish-Gated Residual U-Net. In H. Peng, C. Deng, Z. Wu, and Y. Liu, editors, *Computational Intelligence and Intelligent Systems*, pages 190–204, Singapore, 2019. Springer Singapore. 1, 8, 14, 16
- [16] Y. Liu, Z. Qin, Z. Luo, and H. Wang. Auto-painter: Cartoon Image Generation from Sketch by Using Conditional Generative Adversarial Networks. *CoRR*, abs/1705.01908, 2017. URL <http://arxiv.org/abs/1705.01908>. 8, 9, 12
- [17] Morevna Project. Morevna Project. <https://morevnaproject.org/>, 2024. Acessado em: 8 de novembro de 2024. 3
- [18] NicoNico Seiga. NicoNico Seiga. <http://seiga.nicovideo.jp/>, 2024. Acessado em: 8 de novembro de 2024. 14
- [19] Y. Qu, T. Wong, and P.-A. Heng. Manga colorization. *ACM SIGGRAPH 2006 Papers*, 2006. 5

- [20] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning Transferable Visual Models From Natural Language Supervision. *CoRR*, abs/2103.00020, 2021. 6
- [21] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays. Scribbler: Controlling Deep Image Synthesis with Sketch and Color. *CoRR*, abs/1612.00835, 2016. 9
- [22] M. Shi, J. Zhang, S. Chen, L. Gao, Y. Lai, and F. Zhang. Deep Line Art Video Colorization with a Few References. *CoRR*, abs/2003.10685, 2020. URL <https://arxiv.org/abs/2003.10685>. 11, 15
- [23] M. Shi, J.-Q. Zhang, S.-Y. Chen, L. Gao, Y.-K. Lai, and F.-L. Zhang. Reference-Based Deep Line Art Video Colorization. *IEEE Transactions on Visualization and Computer Graphics*, 29(6):2965–2979, 2023. 10, 11, 15
- [24] L. Stival and H. Pedrini. Survey on Video Colorization: Concepts, Methods and Applications. *Journal of Signal Processing Systems*, 95(6):679–702, May 2023. 2, 5, 12
- [25] R. Street. *Computer Animation: A Whole New World : Groundbreaking Work from Today's Top Animation Studios*. Rockport Publishers, 1998. 1
- [26] D. Sýkora, J. Buriánek, and J. Žára. Unsupervised Colorization of Black-and-White Cartoons. In *3rd International Symposium on Non-Photorealistic Animation and Rendering*, page 121–127, New York, NY, USA, 2004. Association for Computing Machinery. 5, 10
- [27] TaiZan. PaintsChainer Tanpopo, 2016. URL <https://paintschainer.preferred.tech/>. Acessado em: 8 de novembro de 2024. 3, 5, 12, 15
- [28] H. Thasarathan, K. Nazeri, and M. Ebrahimi. Automatic Temporally Coherent Video Colorization. *CoRR*, abs/1904.09527, 2019. URL <http://arxiv.org/abs/1904.09527>. 10, 11, 12, 13, 15
- [29] D. Varga, C. A. Szabó, and T. Szirányi. Automatic Cartoon Colorization Based on Convolutional Neural Network. *15th International Workshop on Content-Based Multimedia Indexing*, 2017. 1, 17
- [30] Walt Disney. Steamboat Willie. Animação em domínio público, 1928. Disponível em plataformas de vídeo público e arquivos online. 3
- [31] H. Winnemöller. XDoG: advanced image stylization with eXtended Difference-of-Gaussians. In *ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, page 147–156, New York, NY, USA, 2011. Association for Computing Machinery. 12
- [32] L. Zhang, Y. Ji, and X. Lin. Style Transfer for Anime Sketches with Enhanced Residual U-net and Auxiliary Classifier GAN, 2017. 7

- [33] L. Zhang, C. Li, T. Wong, Y. Ji, and C. Liu. Two-stage Sketch Colorization. *ACM Transactions on Graphics*, 37:1–14, 2018. 4, 5, 6, 12, 14, 15
- [34] Q. Zhang, B. Wang, W. Wen, H. Li, and J. Liu. Line Art Correlation Matching Network for Automatic Animation Colorization. *CoRR*, abs/2004.06718, 2020. URL <https://arxiv.org/abs/2004.06718>. 10, 11, 13, 15
- [35] R. Zhang, P. Isola, and A. A. Efros. Colorful Image Colorization. *CoRR*, abs/1603.08511, 2016. URL <http://arxiv.org/abs/1603.08511>. 15
- [36] Y. Zhao, D. Ren, Y. Chen, et al. Cartoon Image Processing: A Survey. *International Journal of Computer Vision*, 130:2733–2769, 2022. 4, 8, 9, 10, 12, 15, 16
- [37] C. Zou, H. Mo, R. Du, X. Wu, C. Gao, and H. Fu. LUCSS: language-based user-customized colourization of scene sketches. *CoRR*, abs/1808.10544, 2018. URL <http://arxiv.org/abs/1808.10544>. 7, 14
- [38] C. Zou, Q. Yu, R. Du, H. Mo, Y. Song, T. Xiang, C. Gao, B. Chen, and H. Zhang. SketchyScene: Richly-Annotated Scene Sketches. *CoRR*, abs/1808.02473, 2018. 14
- [39] C. Zou, H. Mo, C. Gao, R. Du, and H. Fu. Language-Based Colorization of Scene Sketches. *ACM Transactions on Graphics*, 38(6), Nov. 2019. 7, 14