

Face Preview Before Plastic Surgery Using Artificial Intelligence Algorithms

Monireh Sadat Fakhraei *

Islamic Azad University, Central Tehran Branch,
Tehran, Iran.

Behzad Khodapanah

Jihad University of Mashhad, Mashhad, Iran.

Abstract

Recent studies have shown significant success in image-to-image translation for two domains. However, existing approaches have limited scalability and robustness in handling more than two domains, as different models must be built independently for each pair of image domains. To overcome this limitation, we propose a model based on generative adversarial networks (GAN), which is a new and scalable approach that can perform image-to-image translation for multiple domains using only a single model. The integrated GAN model architecture allows simultaneous training of multiple datasets with different domains in one network. This leads to the superior quality of images translated by GAN compared to existing models, as well as the new ability to flexibly translate an input image to any desired domain. We experimentally demonstrate the effectiveness of our approach in facial feature transfer and facial expression composite tasks.

Keywords: face recognition, recognition of facial features, translation of images

Received: 11/September/2022

Accepted: 11/November/2022

ISSN: 2980-8936

پیش‌نمایش چهره قبل از عمل جراحی پلاستیک با الگوریتم‌های هوش مصنوعی

منیره السادات فخرائی* | دانشگاه آزاد تهران، واحد مرکز، ایران.

بهزاد خداپناه | جهاد دانشگاهی مشهد، ایران.

چکیده

مطالعات اخیر موفقیت قابل توجهی را در ترجمه‌ی تصویر به تصویر برای دو دامنه نشان داده است. با این حال، رویکردهای موجود، مقیاس‌پذیری و استحکام محدودی را در مدیریت بیش از دو دامنه دارند زیرا مدل‌های مختلف باید به‌طور مستقل برای هر جفت دامنه‌ی تصویر ساخته شوند. برای رفع این محدودیت، ما مدلی بر پایه‌ی شبکه‌های مولد تخصصی (GAN) پیشنهاد می‌کنیم که یک رویکرد جدید و مقیاس‌پذیر است و می‌تواند ترجمه‌ی تصویر به تصویر را برای چندین دامنه‌ی تنها با استفاده از یک مدل واحد انجام دهد. معماری مدل یکپارچه‌ی GAN امکان آموزش هم‌زمان مجموعه داده‌های متعدد با دامنه‌های مختلف در یک شبکه را فراهم می‌کند. این قابلیت، منجر به کیفیت برتر تصاویر ترجمه‌شده توسط GAN در مقایسه با مدل‌های موجود و همچنین، قابلیت جدید ترجمه‌ی انعطاف‌پذیر یک تصویر ورودی به هر دامنه‌ی موردنظر می‌شود. ما به‌طور تجربی، اثربخشی رویکرد خود را در انتقال ویژگی‌های چهره و وظایف ترکیبی بیان چهره نشان می‌دهیم.

کلیدواژه‌ها: شناسایی صورت، تشخیص اجزاء صورت، ترجمه‌ی تصاویر

مقدمه

وظیفه‌ی ترجمه‌ی تصویر به تصویر این است که یک جنبه‌ی خاص از یک تصویر داده‌شده را به جنبه‌ی دیگر تغییر دهد. به‌عنوان مثال، تغییر حالت چهره‌ی یک فرد از خندان به عصبانی. این کار، پیشرفت‌های قابل‌توجهی را پس از معرفی شبکه‌های متخاصم مولد (GAN) با نتایجی از تغییر رنگ مو (Kim et al., 2017)، بازسازی عکس‌ها از حاشیه‌ی تصویر (Isola et al. 2017) و تغییر فصول در تصاویر مناظر (et al., 2017) شاهد بوده است. با توجه به داده‌های آموزشی از دو دامنه‌ی مختلف، این مدل‌ها یاد می‌گیرند که تصاویر را از یک دامنه به دامنه‌ی دیگر ترجمه کنند. چندین مجموعه داده‌ی تصویر دارای برچسب برای ویژگی‌های مختلف چهره نظیر رنگ مو، اندازه‌ی بینی و حالات مرتبط با چهره مثل عصبانیت یا شادابی، در دسترس هستند. به‌عنوان مثال، مجموعه داده‌ی (Liu et al., 2015) CelebA، شامل ۴۰ برچسب مربوط به ویژگی‌های چهره مانند رنگ مو، جنسیت و سن است و مجموعه داده‌ی (Li et al., 2016) RaFD، دارای ۸ برچسب برای حالات چهره مانند شاد، عصبانی و غمگین است. این مجموعه از داده‌ها ما را قادر می‌سازد ترجمه‌ی تصویر به تصویر در چند دامنه را که تصاویر در آن با توجه به ویژگی‌های دامنه‌های متعدد تغییر می‌کند، انجام دهیم. پنج ستون اول در شکل ۱ نشان می‌دهد که چگونه می‌توان یک تصویر CelebA را بر اساس هر یک از چهار دامنه‌ی موی بلوند، جنسیت، سالخوردگی و پوست رنگ‌پریده ترجمه کرد. ما می‌توانیم به آموزش چندین دامنه از مجموعه داده‌های مختلف مانند آموزش مشترک بر پایه‌ی تصاویر CelebA و RaFD برای تغییر حالت چهره‌ی یک تصویر CelebA با استفاده از ویژگی‌های آموخته‌شده توسط آموزش در RaFD (مانند ستون‌های سمت راست شکل ۱) پردازیم. با این حال، مدل‌های موجود در چنین کارهای ترجمه‌ی تصویر به تصویر چند دامنه‌ای، هم ناکارآمد و هم غیر بهینه هستند. نابهینگی آن‌ها از این واقعیت ناشی می‌شود که برای یادگیری همه‌ی نگاشت‌ها در میان k دامنه باید مولدهای $k(k-1)$ آموزش داده شوند. ما از یک برچسب (به‌عنوان مثال، One-Hot Encode) برای نمایش اطلاعات دامنه استفاده می‌کنیم. در طول آموزش، ما به‌طور تصادفی یک برچسب دامنه‌ی هدف تولید نموده و به مدل آموزش می‌دهیم یک تصویر ورودی را به‌طور انعطاف‌پذیر به دامنه‌ی هدف ترجمه کند. با انجام این کار، می‌توانیم برچسب دامنه را کنترل نموده و تصویر را در مرحله‌ی آزمایش به هر دامنه‌ی دلخواه ترجمه کنیم. همچنین، به معرفی یک رویکرد ساده اما مؤثر پرداخته که آموزش مشترک بین دامنه‌های مجموعه داده‌های مختلف را با افزودن یک بردار ماسک به برچسب دامنه امکان‌پذیر می‌سازد. روش پیشنهادی ما می‌تواند نادیده‌گیری برچسب‌های ناشناخته را تضمین نموده و روی برچسب ارائه‌شده توسط یک مجموعه داده‌ی خاص تمرکز کند؛ بنابراین، مدل ما می‌تواند کارهایی مانند ترکیب حالات چهره‌ی تصاویر CelebA را به‌خوبی انجام دهد.

روش تحقیق

هدف ما آموزش یک مدل مولد G است که نگاشت بین چندین دامنه را یاد می‌گیرد. برای رسیدن به این هدف، G را آموزش می‌دهیم تا یک تصویر ورودی x را به یک تصویر خروجی y آموزش دیده بر روی برچسب دامنه‌ی هدف $y = G(x, c)$ ترجمه کند. ما به‌طور تصادفی، برچسب c دامنه‌ی هدف را تولید می‌کنیم تا G یاد بگیرد که تصویر ورودی را به‌صورت انعطاف‌پذیر ترجمه کند. همچنین، یک طبقه‌بندی‌کننده‌ی کمکی را معرفی می‌کنیم که به یک تمایزکننده اجازه می‌دهد چندین دامنه را کنترل کند؛ یعنی تمایزگر ما توزیع‌های احتمال را روی هر دو منبع، ایجاد و برچسب‌های اصلی را تولید می‌کند.

یک مزیت مهم مدل مولد ما این است که هم‌زمان از مجموعه داده‌های متعدد شامل انواع مختلف برچسب‌ها بهره‌مند می‌شود به‌طوری که مدل ما می‌تواند تمام برچسب‌ها را در مرحله‌ی آزمایش کنترل کند. با این حال، یک مسئله

هنگام یادگیری از مجموعه داده‌های متعدد این است که اطلاعات برچسب فقط تا حدی برای هر مجموعه داده شناخته شده است. در مورد CelebA و RaFD، درحالی که مورد اول دارای برچسب‌هایی برای ویژگی‌هایی مانند رنگ مو و جنسیت است، هیچ برچسبی برای حالات چهره مانند شاد و عصبانی ندارد و مورد دوم هم بالعکس. این مسئله برای ما مشکل ساز است زیرا اطلاعات کامل بردار برچسب هنگام بازسازی تصویر ورودی از تصویر ترجمه شده مورد نیاز است.

برای کاهش این مشکل، بردار ماسک را معرفی می‌کنیم که به مدل ما اجازه می‌دهد برچسب‌های نامشخص را نادیده گرفته و بر روی برچسب مشخصاً شناخته شده و ارائه شده توسط یک مجموعه داده‌ی خاص تمرکز کند. در مدل ما، از یک بردار یک‌بعدی n بعدی برای نمایش ماسک استفاده می‌کنیم که n تعداد مجموعه داده‌ها است. همچنین، ما یک نسخه یکپارچه از برچسب را تعریف می‌کنیم. در آزمایش‌های ما، از مجموعه داده‌های CelebA و RaFD استفاده شده است که n در این حالت برابر با دو است.

روش پیشنهادی

به عنوان مدل‌های پایه، دو مدل DIAT (Li et al., 2016) و CycleGAN (Zhu et al., 2017) را انتخاب می‌کنیم که هر دو ترجمه‌ی تصویر به تصویر را بین دو دامنه مختلف انجام می‌دهند. برای مقایسه، ما این مدل‌ها را چندین بار برای هر جفت دو دامنه‌ی مختلف آموزش دادیم. همچنین، ما IcGAN (Perarnau, 2016) را به عنوان یک مدل پایه که می‌تواند انتقال ویژگی را با استفاده از cGAN انجام دهد، انتخاب می‌کنیم.

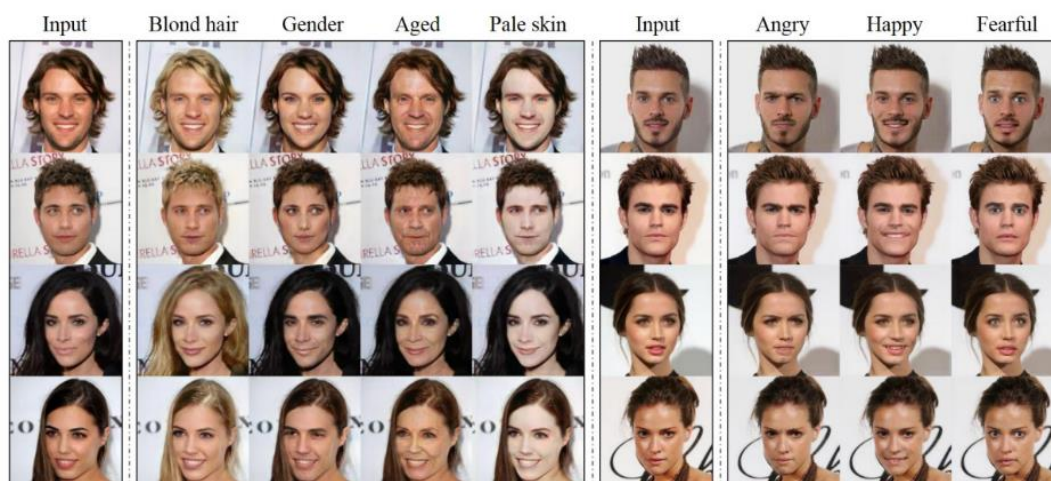
DIAT برای یادگیری نگاشت از $x \in X$ به $y \in Y$ از یک ضرر خصمانه استفاده می‌کند که در آن، x و y به ترتیب تصاویر چهره در دو حوزه‌ی مختلف X و Y هستند. این روش دارای یک اصطلاح منظم سازی در نقشه برداری برای حفظ ویژگی‌های هویت تصویر منبع است (Perarnau et al., 2017).

همچنین، CycleGAN از یک ضرر خصمانه برای یادگیری نگاشت بین دو دامنه‌ی مختلف X و Y استفاده می‌کند. این روش، نگاشت را از طریق تلفات سازگاری چرخه، منظم می‌کند.

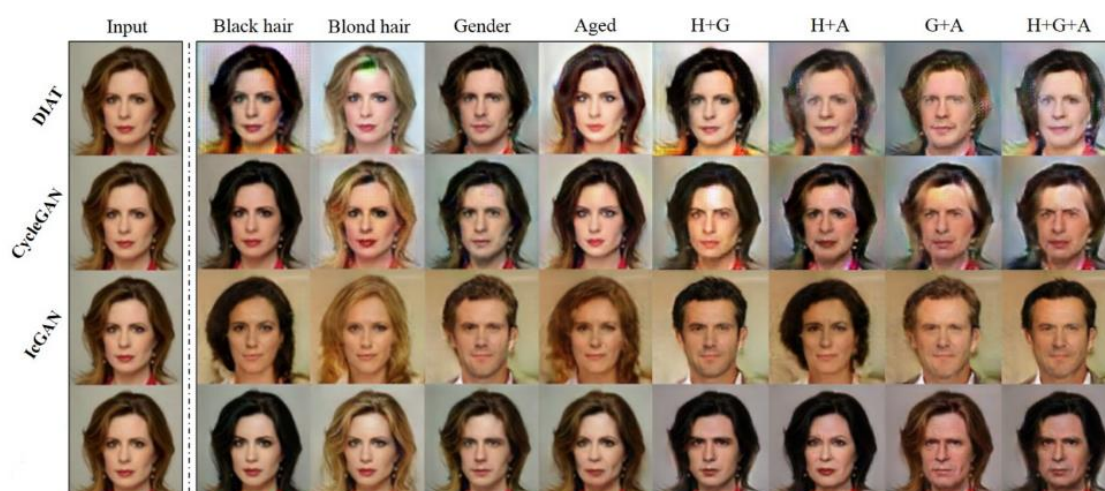
در ابتدا، روش پیشنهادی خود را با مدل‌های پایه در وظایف انتقال تک ویژگی و چند ویژگی مقایسه نموده، مدل‌های دو دامنه مانند DIAT و CycleGAN را چندین بار با در نظر گرفتن همه جفت‌های ارزش ویژگی ممکن آموزش داده (Kim et al., 2017) و در مورد DIAT و CycleGAN، ترجمه‌های چند مرحله‌ای را برای ترکیب چندین ویژگی انجام می‌دهیم (مثلاً انتقال یک ویژگی جنسیتی پس از تغییر رنگ مو) (Hitaj et al., 2017).

مشاهده کردیم که روش ما در مقایسه با مدل‌های رقیب، کیفیت بصری بالاتری از نتایج ترجمه را در داده‌های آزمون ارائه می‌دهد.

علاوه بر این، در مقایسه با IcGAN، مدل ما مزیتی را در حفظ ویژگی هویت چهره‌ی یک ورودی نشان می‌دهد (Lee et al., 2020) چرا که روش ما اطلاعات مکانی را با استفاده از نقشه‌های فعال سازی از لایه‌ی کانولوشن به عنوان نمایش نهفته به جای صرفاً یک بردار پنهان کم بعد مانند IcGAN، حفظ می‌کند.



تصویر ۱. ترجمه‌ی تصاویر صورت به احساسات و ویژگی‌های مختلف



تصویر ۲. خروجی مدل‌های مختلف در ترجمه‌ی تصویر به تصویر

نتیجه‌گیری

در این مقاله، ما مدل مولد متخاصمی را پیشنهاد کردیم که یک مدل ترجمه‌ی تصویر به تصویر مقیاس‌پذیر در بین دامنه‌های مختلف با استفاده از یک مولد واحد و یک مدل متمایزگر است. در مقایسه با روش‌های موجود، مدل ما علاوه بر مزایای مقیاس‌پذیری، تصاویری با کیفیت بصری بالاتر تولید می‌کند که به دلیل قابلیت تعمیم در پشت تنظیمات یادگیری چند وظیفه‌ای است. علاوه بر این، استفاده از بردار ماسک پیشنهادی، بردار مدل ما را قادر می‌سازد تا از مجموعه‌های داده‌ی متعدد با مجموعه‌های مختلف برچسب‌های دامنه استفاده کند؛ بنابراین، تمام برچسب‌های موجود از آن‌ها را مدیریت می‌کند. امیدواریم کار ما به کاربران امکان دهد تا برنامه‌های کاربردی ترجمه‌ی تصویر جالب را در دامنه‌های مختلف توسعه دهند.

منابع

- Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017, October). Deep models under the GAN: information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* (pp. 603-618).
- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

- Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning* (pp. 1857-1865). PMLR..
- Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017, July). Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning* (pp. 1857-1865). PMLR.
- Lee, C. H., Liu, Z., Wu, L., & Luo, P. (2020). Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5549-5558).
- Li, M., Zuo, W., & Zhang, D. (2016). Deep identity-aware transfer of facial attributes. *arXiv preprint arXiv:1610.05586*.
- Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision* (pp. 3730-3738).
- Perarnau, G., Van De Weijer, J., Raducanu, B., & Álvarez, J. M. (2016). Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

