



Computational Intelligence in Electrical Engineering
Vol. 13, No. 4, 2023
Research Paper

Style Transfer for Data Augmentation in Convolutional Neural Networks Applied to Fire Detection

Mahmod Amintoosi

Dept. of Computer Science, Faculty of Mathematics and Computer Science,
Hakim Sabzevari University, Sabzevar, Iran

Abstract:

Adequate training data is essential in all supervised learning methods, including deep learning and machine vision. One of the approaches used to increase the number of training examples in deep learning is the "data augmentation" method. This method involves rotation transformation, transitions, and cropping on training images, which leads to an increase in the number of samples, which are different from training data. In this paper, the "style transfer" algorithm is used to increase the number of training samples. The goal in style transfer is to apply the appearance or visual style of one image to another image. In this paper, this method is used to produce new training examples and as an application, the proposed method is applied to the problem of fire detection. Assuming that the training images recorded during the night are less than the samples taken during the day, by applying a style transfer method, the images of the day are converted into night images and added to the data set as training data. The test results show the efficiency of the proposed data augmentation method. On average, the correct detection rate has increased by 7%.

Keywords: Style Transfer, Data Augmentation, Deep Learning, Fire Detection, Generative Adversarial Networks.



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/isee.2021.124044.1490>

مقاله پژوهشی

انتقال سبک برای افزایش داده‌های آموزشی شبکه‌های کانولوشنی در شناسایی شعله آتش

محمود امین‌طوسی

استادیار، گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه حکیم سبزواری، سبزوار، ایران

m.amintoosi@hsu.ac.ir

چکیده: وجود داده‌های آموزشی کافی، امری اساسی در همه سیستم‌های یادگیری با نظارت و منجمله در حوزه یادگیری عمیق و بینایی ماشین است. یکی از روش‌های استفاده‌شده برای افزایش تعداد نمونه‌های آموزشی در یادگیری عمیق، شیوه «داده‌افزایی» است. این شیوه، متضمن تبدیل‌های دوران، انتقال و برش روی تصاویر آموزشی است که به افزایش تعداد نمونه‌های آموزشی نسبتاً متفاوت از داده‌های اولیه منجر می‌شود. در این نوشتار از الگوریتم «انتقال سبک» مبتنی بر شبکه‌های مولد رقابتی برای افزایش تعداد نمونه‌های آموزشی استفاده شده است. هدف در انتقال سبک، اعمال ظاهر یا سبک بصری یک تصویر روی تصویری دیگر است که جنبه هنری آن بیشتر دیده شده است. در این نوشتار از این شیوه برای تولید نمونه‌های جدید آموزشی استفاده شده و به‌منزله یک کاربرد، روش پیشنهادی بر روی مسئله شناسایی شعله آتش اعمال شده است. با این فرض که تصاویر آموزشی ثبت‌شده در طی شب، کمتر از نمونه‌های اخذشده در روزند، با اعمال یک روش انتقال سبک، تصاویر روز به تصاویر شب، تبدیل و به‌عنوان داده آموزشی به مجموعه دادگان اضافه می‌شوند. نتایج آزمایشات انجام‌شده، کارایی شیوه پیشنهادی را نشان داده است. شیوه پیشنهادی به‌صورت میانگین، ۷ درصد نرخ تشخیص درست را نسبت به استفاده نکردن از آن افزایش داده است.

واژه‌های کلیدی: انتقال سبک، شبکه‌های مولد رقابتی، یادگیری عمیق، شناسایی آتش.

۱- مقدمه

تصویر است. شیوه‌های سنتی بسیاری در ترکیب و ادغام تصاویر ارائه شده‌اند؛ برای مثال، در [۳] از روش‌های هم‌رنگ‌سازی چندباند و تبدیل موجک برای ادغام تصاویر استفاده شده است. در [۴] هدف، انتقال بافت یک تصویر به تصویر دیگر است. این شیوه، بافت یک تصویر مانند پارچه را روی تصویری دیگر اعمال می‌کند. یک جفت تصویر آموزشی (A, A') و تصویر مدنظر (B) برای اعمال سبک به‌عنوان ورودی به آن داده می‌شوند. در یک ساختار هرم گاوسی، به‌ازای هر پنجره از تصویر ورودی، نزدیک‌ترین پنجره به آن از A یافت می‌شود و بافت پنجره متناظر از A' برای تولید B' به کار می‌رود. روش مشابه بعدی، روش فراتفکیک‌پذیری مبتنی بر مثال فریمن و همکاران [۵] بوده است که کلیات کار آن مشابه [۴] است. محدودیت اصلی این شیوه‌ها آن است که فقط از ویژگی‌های سطح پایین

انتقال سبک نقاشی که از دیرباز مورد علاقه هنرمندان بوده است، در حوزه بینایی ماشین و یادگیری عمیق [۱، ۲] نیز توجه زیادی را به خود معطوف کرده است. یک روش انتقال سبک، دو تصویر دریافت می‌کند و تصویر سومی به دست می‌دهد که محتوای کلی آن از تصویر اول و سبک و سیاق (استایل) آن از تصویر دوم حاصل شده باشد. این مسئله تا حدودی مشابه ترکیب (ادغام) تصاویر در پردازش

^۱ تاریخ ارسال مقاله: ۱۴۰۰/۰۴/۲۱

تاریخ پذیرش مقاله: ۱۴۰۰/۰۷/۱۰

نام نویسنده مسئول: محمود امین‌طوسی

نشانی نویسنده مسئول: ایران - سبزواری - دانشگاه حکیم سبزواری - دانشکده ریاضی و علوم کامپیوتر - گروه علوم کامپیوتر

پیشنهادی، نتایج آزمایشات و جمع‌بندی را خواهیم داشت.

(تصویر محتوا)
(تصویر ورودی)



تصاویر تولیدشده

تصاویر استایل (سبک)



(ب)

(الف)

شکل (۱): سطر اول: تصویر روباه، به‌عنوان تصویر محتوا. در سطرهای دوم تا سوم، ستون (الف) نتیجه انتقال سبک عصبی مرجع [۷] با تصویر روباه سطر اول به‌عنوان تصویر محتوا و تصویر سطر مربوطه در ستون (ب) به‌عنوان تصویر استایل است. نمونه‌هایی دیگر در گیت‌هاب نگارنده^۸ مشاهده می‌شوند.

۱-۱- شبکه‌های عصبی کانولوشنی

افزایش قدرت محاسباتی سخت‌افزارها و الگوریتم‌های جدید بهینه‌سازی در حوزه شبکه‌های عصبی موجب توجه مجدد به این شبکه‌ها و به‌ویژه شبکه‌های عصبی کانولوشنی

تصویر استفاده می‌کنند. با گسترش یادگیری عمیق و شبکه‌های عصبی کانولوشنی، شاهد انجام انتقال سبک با این شبکه‌ها - به‌صورتی بسیار کارآمدتر از شیوه‌های سنتی - با نام «انتقال سبک عصبی»^۱ هستیم که با کار تأثیرگذار گتیس^۲ و همکاران [۶] شروع شد. هدف در انتقال سبک، ایجاد تصویر جدیدی براساس دو تصویر ورودی است که محتوای کلی آن از تصویر اول و سبک آن از تصویر دوم (به‌طور مثال، یک اثر هنری) گرفته شده باشد. تصویر اول، تصویر «محتوا»^۳ و دومی، تصویر «استایل» (سبک)^۴ نامیده می‌شوند. شکل ۱ نتیجه اعمال شیوه انتقال سبک عصبی [۷] روی تصویری از یک روباه^۵ و چند سبک مختلف را نشان می‌دهد.

گرچه شاید به جنبه هنری انتقال سبک بیشتر توجه شده است، کاربردهای عملی هم دارد. کاربرد مدنظر در این نوشتار، استفاده از آن برای تولید تصاویر آموزشی بیشتر برای شناسایی شعله آتش است. آنچه در اینجا مدنظر است، شناسایی شیئی (آتش) و یافتن محل آن^۶ در تصویر است که درواقع پیچیده‌تر از طبقه‌بندی تصاویر به دو گروه دارای شعله و بدون شعله است. در روال آموزش یک سیستم و شناسایی اشیاء، به داده‌های آموزشی کافی - وابسته به پیچیدگی مدل - نیاز است. گاهی تعداد نمونه‌ها نسبت به پیچیدگی مدل کم است که به بیش‌برازشی^۷ مدل منجر خواهد شد؛ موردی که در مجموعه دادگان حوزه شناسایی آتش با آن مواجهیم. در بسیاری از مجموعه دادگان حوزه شناسایی آتش، تعداد تصاویر حاوی شعله آتش که در شب اخذ شده‌اند، کمتر از نمونه‌های ثبت‌شده در نور روزند. در این نوشتار با یک روش انتقال سبک، تعدادی از نمونه‌های روز به نمونه‌های شب، تبدیل و در روال آموزش استفاده می‌شوند. نتایج آزمایشات انجام‌شده نشان‌دهنده کارایی شیوه پیشنهادی بوده است.

با توجه به اینکه شیوه پیشنهادی مرتبط با شبکه‌های عصبی کانولوشنی، انتقال سبک، شبکه‌های مولد رقابتی و شناسایی آتش است، در ادامه بخش‌هایی به هر یک از این موارد اختصاص داده خواهد شد. پس از بیان شیوه

مولد را از نمونه‌های اصلی نتواند تفکیک کند.^{۱۶} اگر G و D به ترتیب معرف مولد و متمایزکننده باشند، متمایزکننده به نحوی آموزش داده می‌شود که بتواند نمونه‌های از توزیع واقعی داده‌های p_{data} را از نمونه‌های جعلی با توزیع p_g را از هم تفکیک کند. مولد، نویز $z \sim p_z$ را به عنوان ورودی می‌گیرد و نمونه‌های $G(z)$ از توزیع p_g را تولید می‌کند. تابع هدف GAN در زیر آمده است [۲۶]:

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))] \quad (1)$$

خروجی متمایزکننده، صفر یا یک است. پس لگاریتم آن منهای بی‌نهایت یا صفر است و در نتیجه، بیشینه مقدار عبارت اول صفر خواهد بود. متمایزکننده سعی در شناسایی نمونه‌های تولیدی از مولد به عنوان جعلی دارد (یعنی صفر کردن عبارت دوم)؛ بنابراین، در کل هدف آن بیشینه‌سازی تابع هدف است. همچنین، مولد سعی در فریب متمایزکننده را دارد که به معنی منفی بی‌نهایت شدن عبارت دوم یا کمینه کردن کل عبارت است؛ بنابراین، مولد بهینه، G^* است که رابطه (۲) زیر را برآورده سازد:

$$G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) \quad (2)$$

شبکه اشاره شده در بالا لازم نیست یک شبکه کانولوشنی باشد. شبکه‌های مولد رقابتی با شبکه‌های از نوع کانولوشنی با نام $DCGAN^{17}$ در [۱۴] معرفی و نسخه‌های متنوعی از شبکه‌های مولد رقابتی بر اساس آن ارائه شدند که برخی از آنها در انتقال سبک عصبی استفاده شده‌اند که در بخش بعد به آن پرداخته خواهد شد.

۱-۳- انتقال سبک عصبی

همان‌گونه که پیش از این ذکر شد منظور از انتقال سبک عصبی، اعمال سبک (بافت و رنگ) تصویر استایل روی تصویر محتوای ورودی با استفاده از شبکه‌های عصبی کانولوشنی است. در [۶] برای نخستین بار این ایده مطرح شده بود؛ در یک فرآیند تکراری، خطای بازسازی^{۱۸} بین تصویر حاصله و تصاویر محتوا و سبک را کمینه می‌کند. روال کلی کار به صورت زیر است [۱۳]:

شده است. برخلاف شبکه‌های سنتی چندلایه پرسپترونیک هر نود در یک لایه به تمام نودها (نورون‌های) لایه‌های قبل و بعد از خود متصل است، در شبکه‌های کانولوشنی هر نورون در یک لایه از یک گروه محلی از نورون‌های لایه قبل از خود تأثیر می‌پذیرد [۱]. این شبکه‌ها متضمن چندین لایه از انواع مختلف‌اند که هر دسته از آنها عملکرد مشخصی دارند [۲]. در یک نگاه کلی در یک شبکه کانولوشنی، مشابه یک هرم گاوسی، در هر لایه، تصویر فیلتر و کوچک می‌شود.^۹ لایه‌های کانولوشنی وظیفه استخراج ویژگی‌ها را به عهده دارند؛ لایه‌های ادغام بیشینه، اندازه ورودی را تغییر می‌دهند که ویژگی‌ها در مقیاس‌های مختلف دیده شوند. لایه‌های تمام متصل، کار طبقه‌بندی ویژگی‌های استخراج شده را انجام می‌دهند. مدل‌های VGG^{۱۶} [۸]، DarkNet [۹]، ResNet [۱۰] و AlexNet [۱۱] از جمله مدل‌های مشهور یادگیری عمیق‌اند.

۱-۲- شبکه‌های مولد رقابتی

موضوع اصلی این نوشتار، شبکه‌های مولد رقابتی^{۱۹} (GAN) نیست؛ اما با توجه به اینکه در روش پیشنهادی به کار رفته است، اشاره‌ای به این شبکه‌ها خواهیم داشت. شبکه‌های مولد رقابتی از جمله ابداعات تأثیرگذار دهه اخیر در حوزه علوم کامپیوترند که با مقاله ایان گودفلو^{۲۰}، ژان پورژه-آبادی^{۲۱}، مهدی میرزامؤمن^{۲۲} و همکاران [۱۲] روند رو به رشدی را آغاز کرده و کاربردهای متنوعی منجمله جعل تصویر و ویدئو و انتقال سبک (استایل) پیدا کرده است [۱۳]. هدف اصلی در شبکه‌های مولد رقابتی، تولید نمونه‌های جدید از داده‌های آموزشی است؛ وقتی توزیع داده‌ها در دسترس نیست. در این روش دو شبکه عصبی برای تولید نمونه‌های جدید با یکدیگر به رقابت می‌پردازند؛ این دو شبکه، مولد^{۲۳} و متمایزکننده^{۲۴} نامیده می‌شوند. شبکه مولد تبدیل از یک بردار تصادفی به توزیع مدنظر را فرا می‌گیرد و شبکه متمایزکننده، نمونه‌های تولیدی شبکه مولد را از توزیع واقعی داده‌ها تفکیک می‌کند. آموزش شبکه مولد با این هدف انجام می‌شود که نمونه‌هایی تولید کند که شبکه متمایزکننده را بفریبد و متمایزکننده نمونه‌های تولیدشده

با روش پس انتشار خطا [۱] مشتق‌گیری می‌تواند بر حسب x انجام شود.

تولید سبک (استایل یا شمایل)، ایجاد تصویری است که سبک و سیاق کلی آن مشابه سبک تصویر استایل باشد. استفاده از سبک‌های نقاشی از جمله پرستاده‌ترین تصاویر استایل‌اند که در انتقال سبک به کار برده می‌شوند و هدف، اعمال سبک یک نقاشی روی تصویر ورودی است. به این منظور هم از یک تابع هزینه استفاده می‌شود که تفاوت سبک تصویر استایل و تصویر تولیدی را نشان دهد. بافت و رنگ مشخصه‌های اصلی سبک‌اند که با استفاده از خصوصیات آماری [۱۵] ویژگی‌های استخراج‌شده از لایه‌های اولیه شبکه‌های کانولوشنی محاسبه می‌شوند. ویژگی‌های استخراج‌شده از لایه‌های اولیه شبکه، مشخص‌کننده ویژگی‌های مکانی تصویر، همچون لبه‌ها و اشکال هندسی‌اند که به آنها ساختارهای خرد^{۲۱} هم گفته می‌شود. برای لایه^{۲۱} l ام، میزان تفاوت سبک آن با سبک تصویر استایل به صورت زیر محاسبه می‌شود:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \quad (5)$$

که در آن $G^l \in R^{N_l \times N_l}$ ماتریس گرام حاصل از ضرب داخلی ویژگی‌های لایه^{۲۱} l ام تصویر تولیدی است که به صورت زیر تعریف می‌شود:

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (6)$$

و A^l ماتریس گرام متناظر لایه^{۲۱} l ام تصویر استایل است. نکته اصلی در ماتریس گرام آن است که هم‌رخدادی بافت‌های مختلف در تصویر را مشخص می‌کند؛ برای مثال، اگر بافت صورت و بافت راه‌راه سیاه و زرد با هم رخ دهند، می‌توان انتظار صورت یک ببر را داشت. تابع هزینه مربوط به بخش استایل به صورت مجموع وزن‌دار تابع رابطه^{۲۱} (۵) محاسبه می‌شود:

$$\mathcal{L}_{style}(a, x) = \sum_{l=0}^L w_l E_l \quad (7)$$

که میزان مطابقت هم‌رخدادی یادشده در لایه‌های مختلف دو تصویر - با ضرایب وزنی متفاوت - را محاسبه

۱- ایجاد یک تصویر ورودی تصادفی؛

۲- دادن تصویر به یک شبکه از پیش آموزش دیده^{۲۱}، مانند VGG؛ این شبکه در روال کار آموزش داده نخواهد شد.

۳- محاسبه تابع زیان و مشتق‌گیری براساس پیکسل‌های ورودی؛ بنابراین، فقط ورودی‌ها به‌روزرسانی می‌شوند نه وزن‌های شبکه.

هدف، تغییر تصویر ورودی است؛ به نحوی که محتوای آن از این تصویر و سبک آن از تصویر دیگر باشد. دو مسئله اصلی مرتبط با محتوا و سبک وجود دارد:

ایجاد محتوا: ایجاد تصویری که محتوای آن از تصویر محتوا باشد. محتوای کلی اگر برای مثال، نشان‌دهنده یک منظره است، این محتوا باید حفظ شود؛ اما ظرافت‌ها و بافت و رنگ تصویر (سبک آن) باید از تصویر دیگر اخذ شود. یک تابع زیان بین تصویر محتوای ورودی و تصویر تولیدی باید کمینه شود. برای مقایسه محتوای دو تصویر از ویژگی‌های استخراج‌شده شبکه‌های عصبی کانولوشنی استفاده می‌شود. با توجه به ساختار کلی شبکه‌های کانولوشنی که مثل یک هرم گاوسی به طرف لایه‌های آخر، تصویر کوچک‌تر می‌شود و جزئیات اصلی، حذف و کلیات آن باقی می‌ماند، لایه‌های آخر شبکه نمایانگر محتوای کلی تصویر ورودی‌اند؛ بنابراین، برای لحاظ کردن محتوای تصویر عموماً از لایه‌های آخر شبکه استفاده می‌شود که نمایش‌دهنده ویژگی‌های سطح بالاترند که گاهی به آنها ساختارهای کلان^{۲۲} هم گفته می‌شود [۱۲]. اگر p و x به ترتیب نمایانگر تصویر محتوا و تصویر تولیدشده از شبکه باشند، تابع هزینه این قسمت به صورت زیر تعریف می‌شود:

$$\mathcal{L}_{content}(p, x, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 \quad (3)$$

که در آن $P^l \in R^{N_l \times M_l}$ و $F^l \in R^{N_l \times M_l}$ ویژگی‌های استخراج‌شده در لایه^{۲۱} l ام تصاویر تولیدی و محتوا (p و x) هستند. N_l ، تعداد فیلترها در لایه^{۲۱} l ام و M_l حاصل ضرب درازا و پهنای فیلترهاست. مشتق این تابع بر حسب خروجی تابع فعالیت در لایه^{۲۱} l ام به صورت زیر است:

$$\frac{\partial \mathcal{L}_{content}}{\partial F_{ij}^l} = \begin{cases} (F^l - P^l)_{ij} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \end{cases} \quad (4)$$

می‌کند. مشتق این تابع به صورت زیر است:

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} ((F^l)^T (G^l - A^l))_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0. \end{cases} \quad (8)$$

با داشتن دو تابع هزینه بالا، تابع هدف مسئله انتقال سبک عصبی [۶] به صورت رابطه (۹) زیر خواهد بود:

$$\mathcal{L}_{total}(p, a, x) = \alpha \mathcal{L}_{content}(p, x) + \beta \mathcal{L}_{style}(a, x) \quad (9)$$

که آلفا و بتا ضرایب اهمیت دو عبارت‌اند. در فرآیند انتقال سبک، وزن‌های شبکه اصلی به‌روزرسانی نمی‌شوند؛ بلکه تصویر تولیدی، متغیر مسئله است که در هر دور به‌روزرسانی می‌شود:

$$x = x - \lambda \frac{\partial \mathcal{L}_{total}}{\partial x} \quad (10)$$

این شیوه کلی انتقال سبک در زمینه‌های مختلفی همچون فراتفکیک‌پذیری [۱۶]، ستر معنایی تصویر^{۲۳} [۱۷] و ترجمه تصویر به تصویر^{۲۳} [۱۸] استفاده شده است. در چند سال اخیر پیشرفت‌های زیادی درخصوص شبکه‌های مولد رقابتی [۱۷-۲۱] و روش‌های انتقال سبک مبتنی بر این شبکه‌ها ارائه شده‌اند [۲۱-۲۵]. هدف این نوشتار، توضیح این شیوه‌ها نیست و به دلیل آنکه از یکی از این شیوه‌ها (مرجع [۲۵]) برای داده‌افزایی استفاده شده است، کلیات این شیوه بیان می‌شود.

برخلاف روش‌های معمول انتقال سبک مانند [۷] که یک شبکه کانولوشنی برای یک سبک خاص آموزش می‌بیند، در [۲۵]، یک شبکه برای انتقال استایل چندین سبک آموزش داده شده است. شیوه به کار برده شده در آن، مبتنی بر کارهای قبلی [۱۷، ۲۶] است. فرض اصلی این روش آن بوده است که استایل‌های مشابه، خصوصیات مشترکی دارند؛ برای نمونه، فرض کنید هدف، انتقال سبک نقاشی‌های با سبک امپرسیونیسم^{۲۴} (برداشت‌گرایی) کلود مونه^{۲۵} [۱] روی تصویر محتواس^{۲۴} یا باید برای هر نقاشی یک شبکه جدا آموزش داده شود یا با این فرض که نقاشی‌های با استایل یکسان، خروجی‌های یکسانی در برخی لایه‌های شبکه کانولوشنی خواهند داشت، از میانگین و انحراف معیار این خروجی‌ها برای اصلاح خروجی لایه‌های متناظر تصویر ورودی استفاده کرد. خروجی این لایه‌ها ابتدا نرمال‌سازی و با پارامترهای فراگرفته‌شده از نقاشی‌ها، لغو نرمال‌سازی^{۲۶}

انجام می‌شود. در [۲۶] کارایی این شیوه روی نقاشی‌های مختلف و منجمه نقاشی‌های کلود مونه و با نرمال‌سازی نمونه‌ای شرطی^{۲۷} نشان داده شده است. هدف در این نرمال‌سازی، انتقال خروجی تابع فعالیت یک لایه x به نرمال‌شده آن، z ، وابسته به سبک نقاشی s است که به صورت تبدیل آفین زیر انجام می‌شود:

$$z = \gamma_s \left(\frac{x - \mu}{\sigma} \right) + \beta_s \quad (11)$$

که در آن μ و σ میانگین و انحراف معیار خروجی دسته s و β_s و γ_s پارامترهای آموزش دیده مبتنی بر استایل‌اند. نرمال‌سازی می‌تواند در گروه‌های مختلفی همچون نورون‌ها، ویژگی‌ها و کانال‌ها انجام شود. گروه‌بندی‌های مختلف وزن‌ها در تحقیقات حوزه یادگیری عمیق و منجمه در کاهش تعداد پارامترهای شبکه [۲۷] کاربرد دارد. در [۲۵] مبتنی بر شیوه [۱۷] از نرمال‌سازی در سطح لایه‌ها استفاده شده است. در روش پیشنهادی از این شیوه برای داده‌افزایی در مسئله شناسایی آتش استفاده خواهد شد.

۱-۴- داده‌افزایی در یادگیری عمیق

عموم مدل‌های یادگیری عمیق به لحاظ تعداد زیاد پارامترها، نیازمند حجم فراوان داده‌های آموزشی‌اند. یکی از روش‌های استفاده‌شده برای افزایش تعداد نمونه‌های آموزشی در مواقعی که داده‌های آموزشی به اندازه کافی در دسترس نیستند، شیوه «داده‌افزایی»^{۲۸} است. انعکاس، اضافه کردن نویز، تغییر رنگ، تغییر اندازه و برش تصاویر آموزشی در دسترس از جمله شیوه‌های داده‌افزایی هستند که در مقاله مروری شورت^{۲۹} و خوشگفتار [۲۸] بیان شده‌اند. این شیوه‌های داده‌افزایی در بسترهای مرسوم یادگیری عمیق همچون تنسورفلو^{۳۰} و پای‌تورچ^{۳۱} در دسترس‌اند. به‌تازگی به روش‌های داده‌افزایی مبتنی بر انتقال سبک توجه شده است که عموماً مبتنی بر انتقال سبک تصادفی‌اند؛ برای مثال، در کار^{۳۲} جکسون^{۳۳}، عطاپور و سایرین [۲۹] از سبک‌های متعدد تصادفی برای تولید نمونه‌های جدید استفاده شده است. شیوه ایشان مبتنی بر کار قیاسی و همکاران [۳۰] است. در روش مزبور، تصاویر آموزشی، تحت انتقال سبک تصادفی قرار می‌گیرند و تصاویر جدیدی مبتنی بر داده‌های در دسترس تولید می‌شوند؛ برای مثال، تصویر یک لیوان

در [۳۳] چهار معماری R-FCN, Faster-RCNN [۳۸], SSD [۳۹] و YOLOv3 مقایسه قرار شده‌اند. مطابق نتایج این مقاله، روش YOLOv3 با میانگین دقت ۸۴٫۵ درصد، بالاترین کارایی را در شناسایی آتش داشته است. به‌علاوه همین روش با سرعت پردازش ۲۸ فریم در ثانیه، سریع‌ترین روش در بین روش‌های مقایسه‌شده مبتنی بر شبکه‌های عصبی کانولوشنی بوده است. در مرجع [۳۴] برای روش بهینه‌شده YOLO دقت ۷۶ درصد گزارش شده است. زمان آموزش روی ۱۷۲۰ تصویر و روی CPU، ۱۲ ساعت بوده است. در بخش‌های بعدی این نوشتار از YOLOv3 استفاده خواهد شد.

می‌تواند تحت انتقال سبک قرار گیرد و نمونه‌های جدیدی از تصویر لیوان به وجود آید که در مجموعه آموزشی نبوده‌اند. برای چنین مواردی می‌توان انتظار داشت استفاده از انتقال سبک، به تولید داده‌هایی مشابه نمونه‌های آموزشی منجر می‌شود و کمبود داده‌های آموزشی تا حدودی مرتفع شود. به این ترتیب، امکان مؤثر بودن این شیوه، وابسته به نوع مسئله، تصاویر استایل و روش انتقال سبک خواهد بود.

در ادامه به‌عنوان پیشنهاد این نوشتار، نتیجه استفاده از انتقال سبک در شناسایی شعله آتش در تصویر را خواهیم دید. پیش از آن، مرور مختصری بر موضوع شناسایی آتش می‌شود.

۱-۵- شناسایی آتش

شناسایی زودهنگام حریق نقش بسزایی در اطفای به‌موقع و کاهش خسارات بعدی دارد. مزیت نظارت تصویری نسبت به سایر سنجنده‌های حریق، آن است که نیازی نیست دوربین در محل حریق باشد و گرما یا دود را حس کند، کفایت محل آتش‌سوزی در معرض دید دوربین باشد. موضوع شناسایی آتش از طریق پردازش تصاویر سابقه‌ای حدوداً بیست‌ساله دارد [۳۱]. راهکارهای قدیمی حوزه بینایی ماشین عموماً مناسب شناسایی اشیای صلب‌اند. در سالیان اخیر و با توسعه روش‌های نوین مانند یادگیری عمیق، تشخیص اشیای پدیده‌هایی مانند درخت و آتش که شکل ثابتی ندارند، سهل‌الوصول‌تر شده و تحقیقات زیادی را به خود معطوف کرده است [۳۲-۳۴]. در [۳۳] روش‌های متعددی در شناسایی آتش، بررسی و نتیجه گرفته شده است که روش‌های مبتنی بر شبکه‌های عصبی کانولوشنی از روش‌های سنتی پردازش تصویر کاراترند. در [۳۲] چندین الگوریتم و مدل یادگیری عمیق برای شناسایی و پیدا کردن محل شعله آتش مقایسه شده‌اند: روش Faster-

۲- داده‌افزایی با انتقال سبک عصبی در شناسایی شعله آتش

تهیه داده‌های آموزشی کافی و مناسب برای آموزش یک سیستم شناسایی عموماً وقتگیر و مستلزم هزینه و زمان است. هنگامی که هدف، شناسایی دقیق محل شیء مدنظر در تصویر باشد، اشیای هدف باید در مجموعه دادگان آموزشی علامت‌گذاری شوند که وقت بیشتری نسبت به برچسب‌گذاری تصاویر نیاز دارد. در این بخش به‌عنوان شیوه پیشنهادی در این نوشتار، تأثیر استفاده از انتقال سبک برای افزایش نمونه تصاویر آموزشی در شناسایی آتش بررسی شده است. افزایش تعداد نمونه‌ها به گونه‌ای است که نیازی به علامت‌گذاری تصاویر جدید نخواهد بود. در ادامه از مدل YOLOv3 برای شناسایی آتش، استفاده و تأثیر استفاده از انتقال سبک برای داده‌افزایی در این شیوه بررسی شده است. به این منظور، ابتدا با مجموعه دادگان این حوزه و نحوه علامت‌گذاری تصاویر آموزشی آشنا می‌شویم و سپس مبتنی بر شیوه انتقال سبک عصبی اشاره‌شده در بخش‌های قبل، تعدادی از داده‌های آموزشی اخذشده در نور روز، به شب تبدیل می‌شوند و نتیجه این افزایش داده‌های آموزشی را در مسئله شناسایی آتش خواهیم دید.

۲-۱- مجموعه دادگان شناسایی آتش

پیشنهاد این نوشتار برای استفاده از انتقال سبک به‌منظور داده‌افزایی در شناسایی آتش، تبدیل تصاویر «روز» به تصاویر

RCNN [۳۵] با سه معماری VGG16 [۸]، AlexNet [۱۱] و ResNet [۱۰] اجرا شده و الگوریتم اصلاح‌شده Faster-RCNN با معماری رزنت ۱۰۱، بیشترین کارایی را داشته است. پس از آن، همین الگوریتم، با معماری VGG16 با یک دهم درصد اختلاف در رتبه دوم بوده است. بعد از موارد بالا الگوریتم YOLOv3 [۳۶، ۳۷] نسبت به سایر مدل‌های بررسی‌شده کارایی بیشتری داشته است.

گروه حاوی آتش یا غیر آن مناسباند؛ اما برای شناسایی محل آتش به علامت‌گذاری ناحیه آتش نیاز دارند. برخی از مجموعه دادگان فقط شامل چندین فریم از صحنه‌های تقریباً ثابت حاوی آتش‌اند.

در [۴۱] یک معماری سبک شناسایی آتش برای کار روی رسیپری پای پیشنهاد و پیاده‌سازی^{۳۹} شده است. مجموعه دادگان ارائه شده شامل انواع مختلفی از تصاویر حریق در اختیار عموم قرار داده شدند. برای هر تصویر ناحیه آتش به صورت دستی مشخص شده است. در این نوشتار، از بخشی از این مجموعه داده استفاده شده که از گیت‌هاب مقاله در دسترس است.

مشخص کردن دستی ناحیه اشیای مدنظر در هر تصویر با ابزارهای مختلف مانند `labelImg`^{۴۰} انجام می‌شود. به کادر مشخص‌کننده ناحیه شیء، جعبه محدودکننده^{۴۱} گفته می‌شود. شکل ۲ دو نمونه از تصاویر مجموعه داده فوق به همراه علامت‌گذاری^{۴۲} جعبه‌های محدودکننده نواحی آتش را نشان می‌دهد. این کادرها به عنوان علامات مرجع^{۴۳} مشخص‌کننده محل درست شیء، بعداً در روال ارزیابی استفاده خواهند شد.



(ب) علامت‌گذاری نواحی دارای شعله آتش با کادر مستطیلی.

(الف) دو تصویر آموزشی از مجموعه دادگان [۴۱]

شکل (۲): نمونه تصاویری از مجموعه دادگان استفاده شده و شیوه علامت‌گذاری دستی نواحی آتش.

همان‌گونه که مشاهده می‌شود در یک تصویر ممکن است چند مورد از شیء مدنظر وجود داشته باشد. مشخصات نواحی هر تصویر در قالب یک فایل xml ذخیره می‌شود.

«شب» است. با این فرض که تصاویر آموزشی ثبت شده در طی روز، بیشتر از نمونه‌های اخذ شده در شب‌اند، با اعمال یک روش انتقال سبک، می‌توان کمبود نمونه‌های اخذ شده در شب را برطرف کرد و افزایش کارایی سیستم شناسایی را انتظار داشت.

به جای بررسی یک سیستم شناسایی حریق همه‌منظوره که در همه زمان‌ها و مکان‌ها کار کند، روی موردی تمرکز خواهیم کرد که مناسب با ایده این نوشتار باشد. به این منظور از تصاویری شامل شعله آتش استفاده خواهد شد که در فضای باز، گرفته و عمدتاً شامل حریق خودرو در بزرگراه‌ها یا خیابان می‌شوند؛ زیرا مایلیم انتقال سبک با تصاویر شب در دسترس روی مجموعه دادگان استفاده شده امکان‌پذیر باشد. فرض آموزش و تست یک سیستم روی مجموعه داده‌ای خاص، فرضی غیرمنطقی نیست. طراحی یک سیستم شناسایی خاص، کاراتر از یک سیستم شناسایی همه‌منظوره است؛ برای مثال، در همین مورد آتش‌سوزی، با توجه به تفاوت‌های فراوان بین محیط خانه و بیرون، نوع ماده در حال سوخت، رنگ و حجم شعله، حریق در یک پالایشگاه یا کارخانه یا ایستگاه انتقال گاز، قاعدتاً یک سیستم شناسایی آتش تخصصی بهتر می‌تواند جوابگو باشد. در این بخش، به مجموعه داده‌ای نیاز است که تصاویر آن تقریباً یک سبک و سیاق داشته باشند، محل شعله در تصاویر علامت‌گذاری شده و در دسترس عموم باشد.

در بخش‌های پیش به چند روش شناسایی آتش اشاره شد. در [۳۳] گفته شد از بین چند مدل بررسی شده، YOLOv3 بیشترین کارایی را داشته است. در این مرجع به منظور آموزش سیستم، ۲۹۱۸۰ تصویر از منابع متعدد گردآوری و برچسب زده شده‌اند. از میان این تعداد، ۱۳۴۰۰ تصویر شامل آتش/دود بوده‌اند. مجموعه داده استفاده شده در مقاله فوق در دسترس عموم نبود؛ در مقاله مذکور به چندین مجموعه داده دیگر در حوزه شناسایی آتش اشاره شده بود که با توجه به فرض پاراگراف قبلی، مناسب این نوشتار نیستند. مجموعه داده‌های دانشگاه کیم‌یانگ^{۴۴}، کورسیکا^{۴۵}، بیلکنت^{۴۶} و دانشگاه اسپلیت^{۴۷}، مربوط به آتش‌سوزی در جنگل است یا شامل برچسب نواحی آتش نبودند. در مرجع [۴۰] از سه گیگابایت تصویر^{۴۸} برای آموزش استفاده شده است. چنین مجموعه داده‌هایی برای طبقه‌بندی تصاویر به دو

۲-۲- اعمال انتقال سبک روی مجموعه داده

از بین ۳۰۵ تصویر انتخابی، ۲۵ تصویر در شب و سایر تصاویر در روز گرفته شده‌اند. ۸۵ تصویر روز، انتخاب و با استفاده از انتقال سبک عصبی، به تصویر شب تبدیل شدند. شکل ۳ دو تصویر نمونه را نشان می‌دهد که تحت انتقال سبک عصبی با دو روش قرار گرفته‌اند. سطرهای با شماره ۱ تا ۴ در شکل ۳ به ترتیب تصویر محتوای ورودی، تصویر استایل، نتیجه انتقال سبک با روش [۷] و نتیجه انتقال سبک با روش [۲۵] روی تصاویر ستون‌های (الف) و (ب) را نشان می‌دهند. تصاویر استایل استفاده‌شده، از مجموعه دادگان برکلی^{۴۴} است که در مرجع [۱۸] استفاده شده است. حجم

کل این مجموع دادگان ۹ گیگابایت است^{۴۵} که بخشی از آنها برای استفاده در این نوشتار در گیت‌هاب نگارنده^{۴۶} قرار داده شده است. برای هر تصویر روز، یک تصویر شب، تصویر استایل آن در نظر گرفته می‌شود و عمل انتقال سبک روی آن به منظور تبدیل تصویر روز به تصویر شب انجام می‌شود. تصویر شب به صورت تصادفی از مجموعه تصاویر شب انتخاب شده است. انتقال سبک با دو شیوه [۷، ۲۵] انجام و نتایج در شکل ۳ نشان داده شده است. با توجه به شکل ۳، روش انتقال سبک [۲۵] نتایج بهتری نسبت به روش [۷] در این کاربرد تولید کرده است.



(ب)



(الف)

۱. تصویر
محتوای
ورودی۲. تصویر
استایل۳. انتقال
سبک با
روش [۷]۴. انتقال
سبک با
روش [۲۵]

شکل (۳): نتیجه اعمال دو روش انتقال سبک عصبی روی نمونه تصاویری از مجموعه دادگان استفاده‌شده.

یادگیری عمیق اند که وابسته به کاربرد مدنظر همه یا برخی از آنها روی تصاویر آموزشی اعمال می‌شوند؛ برای مثال، اگر کاربرد مدنظر شناسایی عابر پیاده باشد، انعکاس افقی تصویر مناسب است؛ اما انعکاس در راستای محور yها راهگشا نخواهد بود. به منظور ارزیابی کارایی سیستم پیشنهادی، چهار روش زیر در شناسایی آتش بررسی خواهند شد:

- ا) بدون هیچ‌گونه داده‌افزایی (Aug-None)
 ب) داده‌افزایی از نوع تغییر اندازه و برش (Aug-Scale)
 ج) داده‌افزایی از نوع انتقال سبک (Aug-D2N)
 د) داده‌افزایی از نوع انتقال سبک + تغییر اندازه و برش^{۴۸} (Aug-D2N-Scale)

چارچوب کلی کار، مطابق توضیحات بخش‌های قبلی در شکل ۴ نشان داده شده است که بخش اول آن بسته به هر یک از چهار روش یادشده تغییر خواهد کرد. همه برنامه‌ها روی سرورهای گوگل کولب با مشخصات ذکرشده در جدول ۱ اجرا شده‌اند.



شکل (۴): چارچوب کلی روش پیشنهادی

جدول (۱): مشخصات دستگاه

GPU	RAM	CPU
Tesla T4, 15 GB	13 GB	Intel Xeon CPU @ 2.20GHz

یک مدل از قبل آموزش دیده YOLOv3 روی تصاویر طبیعی به عنوان مدل اولیه آموزش انتخاب شد. با توجه به اینکه معماری این مدل متفاوت با مدل‌های کراس^{۴۹} استفاده‌شده برای آموزش مدل است، ساختار این مدل به

برای هر دو روش از مدل‌های از پیش آموزش دیده نگارندگان آنها در بستر پای‌تورچ استفاده شده است. فایل‌های مدل‌های آموزش‌دیده این شیوه که در انتقال سبک به کار می‌روند، حدود ۵۰ مگابایت است؛ در حالی که فایل مدل مولد روش [۲۵] حدود یک و نیم گیگابایت است. با توجه به اینکه روش اخیر روی تبدیل تصاویر روز به شب آموزش دیده، نتایج بهتری به دست داده است؛ بنابراین، در ادامه از این شیوه برای تبدیل تصاویر روز به شب استفاده شده است.

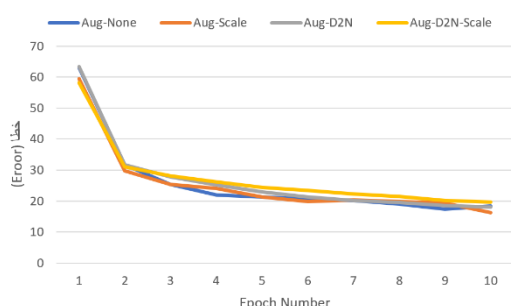
۲-۳- شناسایی آتش در تصویر

همان‌گونه که پیش‌تر ذکر شد روش‌های مبتنی بر شبکه‌های عصبی کانولوشنی در مقایسه با روش‌های مرسوم پردازش تصویر، کارایی بهتری در شناسایی آتش در تصویر دارند [۳۳]. در ادامه برای آموزش و شناسایی آتش در تصاویر از کتابخانه منبع باز ImageAI^{۴۷} استفاده خواهد شد. مزیتی که شیوه انتقال سبک در تبدیل تصاویر روز به شب در این خصوص دارد، آن است که نواحی از قبل مشخص شده به عنوان حریق (شکل ۲) قابلیت استفاده را برای تصاویر تولیدی با انتقال سبک دارند (دو سطر آخر شکل ۳). به علاوه، این نکته که سایر نواحی تصویر شامل آتش نیستند، نورهای تولیدشده در عمل انتقال سبک، به عنوان نواحی که آتش نیستند، در فرآیند آموزش در نظر گرفته خواهند شد. برای مثال، با دقت در تصاویر سطر آخر شکل ۳، ملاحظه می‌شود به جز نور چراغ‌های ماشین‌ها و نور چراغ گردان ماشین آتش‌نشانی در ستون (الف)، نورهای مصنوعی دیگری منبعث از تصاویر استایل به این تصاویر اضافه شده‌اند که چون جزو نواحی آتش نیستند، در فرآیند آموزش به عنوان آتش نباید تلقی شوند و مدل آموزشی باید در برابر چنین نورهایی که درواقع آتش کاذب‌اند، مقاوم‌تر شود.

۲-۴- نتایج اجرا و ارزیابی

همان‌گونه که پیش‌تر ذکر شد انعکاس، دوران، تغییر اندازه و برش از جمله روش‌های معمول داده‌افزایی در حوزه

مشاهده می‌شود تغییرات تابع هزینه در چهار مدل تقریباً مشابه هم بوده‌اند.



شکل (۵): نمودار کاهش مقدار تابع هزینه در حین آموزش شبکه برای چهار مدل بررسی شده.

برای ارزیابی دقیق‌تر، مدل انتخابی هر روش با معیارهای مرسوم این حوزه مقایسه خواهند شد. محاسبه دقت در سیستم‌های شناسایی محل شیئی مبتنی بر جعبه محدودکننده شیئی است. ابتدا تعریف چند معیار، مرور و سپس «میانگین دقت» ذکر شده‌اند.

اشتراک به اجتماع (IoU^{66}): معیاری است که متناسب با میزان هم‌پوشانی جعبه محدودکننده مرجع^{۶۷} و جعبه محدودکننده حاصل از مدل است و به‌صورت زیر محاسبه می‌شود:

$$IoU = \text{Area of intersection} / \text{Area of union} \quad (12)$$

اگر b_{gt} مشخص‌کننده کادر مرجع شیئی و b_p مشخص‌کننده کادر پیش‌بینی شده باشد، IoU به‌صورت زیر بیان می‌شود:

$$IoU = \text{Area}(b_{gt} \cap b_p) / \text{Area}(b_{gt} \cup b_p) \quad (13)$$

اگر IoU از حد آستانه خاصی بیشتر باشد، مورد شناسایی شده به‌عنوان تشخیص درست منظور خواهد شد. حد آستانه، عددی عموماً در بازه ۰٫۰۵ تا ۰٫۹۵ است که هر چه بیشتر باشد، در قبول ناحیه شناسایی شده به‌عنوان درست، سخت‌گیرانه‌تر عمل می‌کند. وابسته به مقدار این حد آستانه میزان دقت برآوردشده متفاوت خواهد شد؛ بنابراین، میانگین دقت براساس سطح زیر منحنی درستی - نرخ یادآوری^{۶۸} محاسبه می‌شود. به این ترتیب که در سطوح

قابل خواندن در کراس تغییر یافت.

در هر روش، ابتدا مدل روی تصاویر مجموعه دادگان انتخابی آموزش داده می‌شود. پس از ده اپک^{۶۹} مدل با کمترین خطا روی داده‌های اعتبارسنجی^{۷۱} به‌عنوان مدل نهایی انتخاب می‌شود. روال کلی، همان فرم نمایش داده شده در شکل ۴ است. تعداد تصاویر اعتبارسنجی، ۶۰ نمونه بوده است. برنامه مربوط به آموزش شبکه از گیت‌هاب نگارنده در دسترس و اجرایشدن است^{۷۲}.

در روش (آ)، آموزش روی ۳۰۵ تصویر آموزشی و بدون هیچ داده‌افزایی صورت پذیرفته است (با نام Aug-None). روش (ب)، عیناً مشابه روش (آ) بوده است؛ با این تفاوت که در ۳۰ درصد مواقع، تصویر آموزشی از داده‌افزایی از نوع تغییر اندازه و برش تأثیر گرفته است. چون ده اپک برنامه آموزش داده می‌شود، هر تصویر، هم با اندازه اصلی و هم دست‌کم یک بار با تغییر اندازه و برش در آموزش مشارکت داشته است. این شیوه Aug-Scale نامیده شده است. پارامترهای اصلی برنامه در جدول ۲ آمده‌اند.

جدول (۲): پارامترهای اجرا

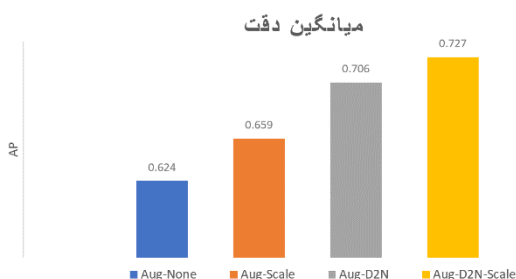
پارامتر	مقدار
بستر برنامه‌نویسی	تنسورفلو
تعداد تصاویر آموزشی برای روش‌های (آ) و (ب)	۳۰۵
تعداد تصاویر روز به شب شده در روش‌های (ج) و (د)	۸۵
تعداد اپک آموزشی	۱۰
احتمال اعمال داده‌افزایی با روش Scale&Crop	٪۳۰
بازه تغییر اندازه	[۰٫۲۵ و ۲]
نرخ یادگیری	۰٫۰۰۰۰۱
حد آستانه IoU	۰٫۲

در روش (ج) با نام Aug-D2N^{۷۳}، ۸۵ تصویر از ۳۰۵ تصویر (حدود ۳۰ درصد) تحت انتقال سبک عصبی قرار گرفتند و به مجموعه دادگان آموزشی اضافه شدند. در روش (د) با نام Aug-D2N-Scale، مشابه شیوه (ب) اجازه تغییر اندازه و برش به تصاویر آموزشی (کل ۳۹۰ تصویر) داده می‌شود.

روند کاهش مقدار تابع هزینه^{۷۴} شبکه برای چهار مدل فوق در شکل ۵ نمایش داده شده است. همان‌گونه که

دقت داشته است که نشان‌دهنده کارایی روش پیشنهادی در ترکیب با روش تغییر اندازه است.

در ادامه برخی از خروجی‌های دو روش (ب) و (د) یعنی Aug-Scale و Aug-D2N-Scale ملاحظه می‌شود. با معیار میانگین دقت یادشده، روش دوم حدود ۷ درصد افزایش دقت داشته است.



شکل (۶): مقایسه دقت نهایی دو مدل در شناسایی ناحیه آتش، مبتنی بر معیار میانگین دقت (Average Precision).

شکل ۷ نتیجه اجرای دو مدل روی چند تصویر نمونه را نشان می‌دهد. ستون (الف) نتایج مدل Aug-Scale و ستون (ب) نتایج مدل پیشنهادی است که تعدادی از تصاویر آموزشی روز با روش انتقال سبک عصبی به شب تبدیل شده‌اند. همه پارامترهای برنامه و نمونه‌های اعتبارسنجی و آزمون، در هر دو مدل یکی و بدون تغییر بوده‌اند. در دو سطر اول، روش پیشنهادی عملکرد بهتری داشته است. در دو سطر بعد، روش اول، در دو سطر ۵ و ۶، هر دو روش شعله را شناسایی کرده‌اند و در دو سطر آخر هر دو ناموفق بوده‌اند. یک نکته شایان توجه، شناسایی بهتر روش پیشنهادی در سطر پنجم است. بدون داده‌افزایی پیشنهادی (ستون اول) چراغ‌های گردان خودرو به اشتباه، شعله آتش شناسایی شده‌اند.

در مجموعه تصاویر آزمون، ۱۶ تصویر شب وجود داشته است. روش پیشنهادی توانسته است در ۸ مورد، شعله آتش را شناسایی کند؛ در حالی که بدون روش پیشنهادی، فقط دو نمونه شناسایی درست داشته است. کارایی بهتر روش پیشنهادی در تصاویر شب به دلیل افزایش نمونه‌های شب با روش انتقال سبک روز به شب است. شکل ۸ چهار نمونه از تصاویر شب را نشان می‌دهد که روش پیشنهادی موفق بوده است.

مختلف IoU در بازه $[0,1]$ تعداد مثبت‌های صادق^۹، مثبت‌های کاذب^{۱۰} و منفی‌های کاذب^{۱۱} و براساس آنها درستی و نرخ یادآوری محاسبه می‌شوند:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{درستی:} \quad (14)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{نرخ یادآوری:}$$

سطح زیر منحنی درستی - نرخ یادآوری، میانگین دقت (AP) است که تعریف عمومی آن به صورت زیر است:

$$AP = \int_0^1 p(r) dr \quad (15)$$

که منظور از p و r در رابطه بالا همان precision و recall یادشده‌اند.

با محاسبه معیارهای درستی و نرخ یادآوری برای چهار شیوه مورد بحث، میانگین دقت برای هر دو مدل مبتنی بر روش پیشنهادی داده‌افزایی با انتقال سبک، بیشتر از دو روش دیگر بوده است (شکل ۶). افزایش دقت روش داده‌افزایی تغییر اندازه و برش (Aug-Scale) نسبت به روش بدون داده‌افزایی Aug-None، کمتر از ۴ درصد (۰٫۳۵) بوده است؛ در حالی که افزایش دقت روش پیشنهادی Aug-D2N نسبت به روش بدون داده‌افزایی Aug-None بیش از ۸ درصد بوده است؛ به این معنی که شیوه پیشنهادی به‌عنوان یک روش داده‌افزایی در مسئله شناسایی آتش مؤثرتر از روش تغییر اندازه و برش است.

به این نکته باید دقت داشت که صرف افزایش دقت شیوه پیشنهادی نسبت به روش بدون داده‌افزایی، حتی اگر میزان افزایش دقت آن از میزان افزایش دقت روش تغییر اندازه هم کمتر می‌بود، باز هم شیوه پیشنهادی به‌عنوان یک روش داده‌افزایی پذیرفتنی بود؛ زیرا روش‌های داده‌افزایی می‌توانند مستقل از هم به کار برده شوند. نوع داده‌افزایی که روش پیشنهادی می‌تواند انجام دهد، انجام‌پذیر با روش تغییر اندازه نیست و برعکس.

به‌منظور بررسی تأثیر ترکیب دو روش پیشنهادی و روش تغییر اندازه، روش تغییر اندازه روی ۳۹۰ تصویر حاصل از روش پیشنهادی اعمال شده است (با همان احتمال پیشین). این روش با نام Aug-D2N-Scale در مقایسه با روش مبتنی بر تغییر اندازه حدود ۷ درصد افزایش



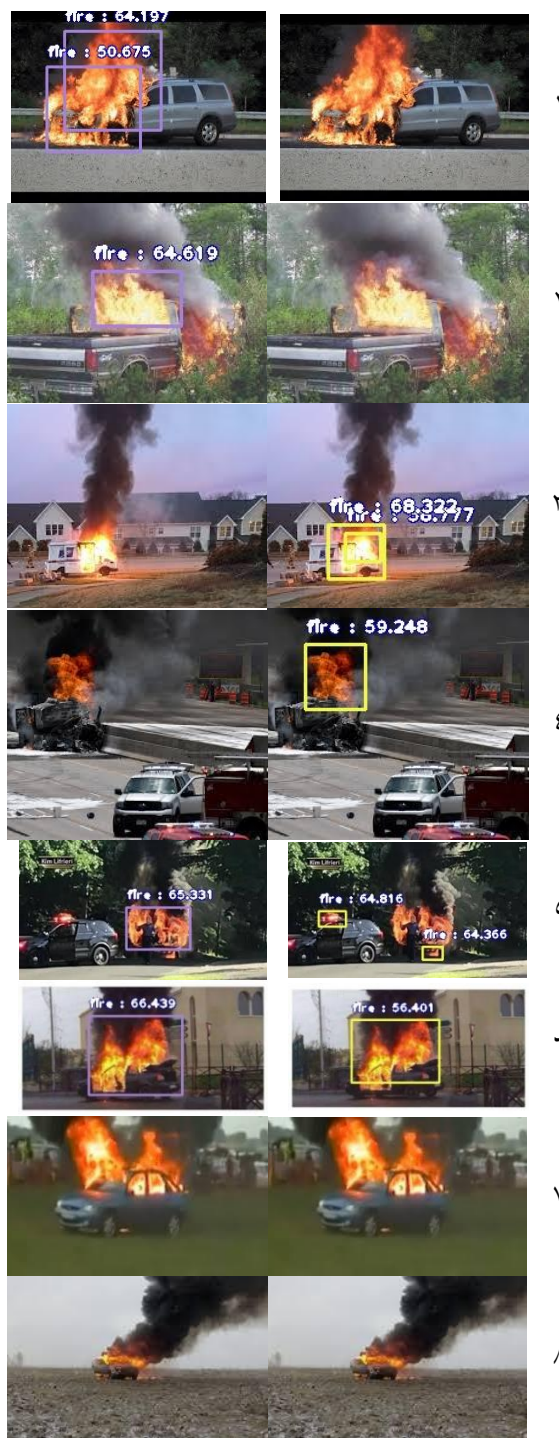
(الف) روش Aug-Scale (ب) روش پیشنهادی
(Aug-D2N-Scale)

شکل (۸): نتایج شناسایی آتش در شب توسط دو روش

Aug-D2N-Scale و Aug-Scale

زمان اجرا

با افزایش ۸۵ تصویر به ۳۰۵ تصویر مجموعه دادگان، حجم داده‌ها ۲۸ درصد اضافه شد که طبقاً زمان آموزش نیز زیاد خواهد شد. زمان آموزش هر دو مدل (آ) و (ب) روی ۳۰۵ تصویر در ۱۰ اپک، حدود ۳۸ دقیقه و زمان اجرای مدل‌های پیشنهادی (ج) و (د) روی $305+85=390$ تصویر (۲۸ درصد افزایش تعداد تصاویر)، حدود ۴۸ دقیقه بوده است که معادل ۲۶ درصد افزایش زمان آموزش است (شکل ۹)؛ اما ساختار و حجم هر دو مدل یکی بوده (۲۴۷ مگابایت) و پس از آموزش، زمان شناسایی یکسان است (۷,۵ ثانیه برای ۶۰ تصویر) و به‌صورت معمول، آنچه مهم است زمان اجرای مدل است که هر دو سرعت یکسانی در شناسایی محل حریق دارند (۸ تصویر در ثانیه).



(الف) روش Aug-Scale (ب) روش پیشنهادی
(Aug-D2N-Scale)

شکل (۷): نتایج شناسایی آتش با دو روش Aug-Scale و

Aug-D2N-Scale

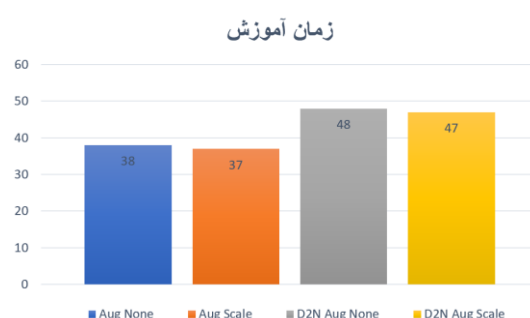
تصاویر آموزشی از جمله روش های مرسوم داده افزایی در یادگیری عمیق است. به تازگی روش های انتقال سبک هم برای افزایش تعداد داده ها به کار گرفته شده اند. در این مقاله دو روش انتقال سبک عصبی برای تبدیل تصاویر روز به شب به کار گرفته شدند. روشی که نتایج بصری بهتری تولید کرده بود به عنوان شیوه داده افزایی در شناسایی آتش به کار گرفته شد. نتایج آزمایشات انجام شده نشان دادند در این کاربرد خاص شیوه پیشنهادی به افزایش دقت شناسایی منجر می شود.

روش انتقال سبک در برابر روش های معمول داده افزایی در یادگیری عمیق (همچون برش تصویر، تبدیلات هندسی و تغییر رنگ) نیست؛ بنابراین، بحث مقایسه این دو شیوه مطرح نیست. شیوه های مرسوم داده افزایی روی تصاویر اضافه شده با انتقال سبک هم اعمال می شود؛ با این حال، شیوه پیشنهادی با روش داده افزایی تغییر اندازه و برش مقایسه شد که نتایج بهتری از آن به دست داد.

هدف در نوشتار حاضر، ایجاد بهترین مدل برای شناسایی آتش نبوده است؛ به همین دلیل روی مواردی همچون تعداد تصاویر آموزشی، تعداد اپک آموزشی، پارامترهای آموزش مدل، نرخ یادگیری، توابع فعالیت و الگوریتم های بهینه سازی مطالعه خاصی انجام نشده است.

برای مقایسه کارایی روش داده افزایی مبتنی بر تغییر سبک روز به شب در شناسایی آتش، تمام تنظیمات و پارامترهای آموزشی مدل های مختلف بررسی شده مانند هم در نظر گرفته شد و نشان داده شد افزایش نمونه ها با روش انتقال سبک عصبی [۲۵] کارایی سیستم شناسایی آتش را ارتقا می دهد. یک مزیت روش پیشنهادی آن است که می توان از همان جعبه های مشخص کننده نواحی آتش در تصاویر آموزشی روز، برای متناظر شب آنها هم استفاده کرد؛ بدون آنکه نیاز به علامت گذاری مجدد این تصاویر باشد. به علاوه با توجه به اینکه تصاویر استایل شب استفاده شده شامل نورهای چراغ و نور لامپ ها بودند و نورهای مشابه تصنعی در تصاویر روز به شب شده ایجاد می شود، به صورت ضمنی روال آموزش، چنین نورهایی را از آتش تفکیک خواهد کرد.

شیوه انتقال سبک بررسی شده در این نوشتار می تواند در



شکل (۹): زمان آموزش بر حسب دقیقه

انتقال سبک نامناسب

در نتیجه اعمال انتقال سبک روی برخی از تصاویر، ناحیه شعله در تصویر حاصله تا حدود زیادی رؤیت ناپذیر شده بود. شکل ۱۰ دو نمونه از چنین مواردی را نشان می دهد. در روال آموزش اشاره شده در بخش های قبل، از تمام تصاویر روز به شب شده و منجمله این دو تصویر استفاده شده بود؛ اما در حالت کلی باید یا چنین تصاویر تولید شده ای به صورت دستی از مجموعه دادگان حذف شوند یا راهکاری برای حفظ ناحیه شعله آتش در هنگام انتقال سبک اندیشیده شود که از جمله کارهای آتی می تواند باشد.



تصویر محتوا تصویر استایل نتیجه انتقال سبک
شکل (۱۰): برخی از تبدیل های نامناسب روز به شب که در تصویر حاصل (ستون سوم) ناحیه آتش تقریباً محو شده است.

۳- جمع بندی

با ظهور شبکه های عصبی کانولوشنی در حوزه شناسایی اشیا در تصاویر جهش بزرگی رخ داده است. بسیاری از مدل های شبکه های عصبی کانولوشنی شامل میلیون ها پارامتر بوده که نیازمند حجم زیادی داده آموزشی اند. افزایش تعداد تصاویر آموزشی از طریق اعمال تبدیلات هندسی روی

- 2002, doi: 10.1109/38.988747.
- [6] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 241
- [7] X. Li, S. Liu, J. Kautz, and M. Yang, "Learning Linear Transformations for Fast Image and Video Style Transfer," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15-20, 2019, pp. 3804-3812.
- [8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015, in *International Conference on Learning Representations*.
- [9] J. Redmon and A. Farhadi, "YOLO9000 :Better, Faster, Stronger," *CVPR*, 2017.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 770-778.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger Eds.: Curran Associates, Inc., 2012, pp. 1097-1105.
- [12] I. J. Goodfellow *et al.*, "Generative adversarial nets," presented at the Proceedings of the 27th International Conference on Neural Information Processing Systems – Vol. 2, Montreal, Canada, 2014.
- [13] F. Chollet, *Deep Learning with Python*. Manning Publications Co., 2017.
- [14] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," 2016. [Online]. Available: <http://arxiv.org/abs/1511.06434>.
- [15] J. Portilla and E. P. Simoncelli, "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients," *International Journal of Computer Vision*, Vol. 40, No. 1, pp. 49-70, 2000.
- [16] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *14th European Conference on Computer Vision*, B. Leibe, J. Matas, N. Sebe, and M. Welling Eds., Lecture Notes in Computer Science. Amsterdam, 2016. Springer.
- [17] T. Park, M.-Y. Liu, T. Wang, and J.-Y. Zhu, "Semantic Image Synthesis With Spatially-Adaptive Normalization," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2332-2341, 2019.
- [18] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 22-29 Oct. 2017 2017, pp. 2868-2876.
- [19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks,"

دیگر مسائل شناسایی اشیا در تصاویر هم استفاده شود که فاقد داده‌های آموزشی کافی از برخی دسته‌ها هستند؛ برای مثال، ممکن است در یک سیستم شناسایی برگ یا گیاه مریض از سالم، تعداد نمونه‌های مریض کمتر باشند که می‌توان با یک روش انتقال سبک، تصاویر سالم را به مریض تبدیل کرد و تعداد نمونه‌های آموزشی را افزایش داد. کاربردهای مشابه دیگری همچون شناسایی کاشی معیوب از کاشی سالم و شناسایی خودروی آسیب‌دیده از خودروی سالم هم می‌توان برای ایده این نوشتار در نظر گرفت. ازجمله کارهای آتی می‌توان به بهبود عملکرد روش انتقال سبک اشاره کرد؛ به نحوی که ناحیه شعله آتش در تصاویر روز، کمترین تأثیر را بپذیرند. در حال حاضر ناحیه شعله در تصاویر تولیدشده با انتقال سبک، مقداری تصنعی‌اند. استفاده از فضاهای رنگی دیگر، پیدا کردن بهترین پارامترهای شبکه و استفاده از تصاویر آموزشی بیشتر، مواردی‌اند که به‌منظور حصول به یک سیستم شناسایی آتش بهینه به بررسی بیشتر نیاز دارد که موضوع این نوشتار نیست.

سپاسگزاری

از داوران گرامی که با نظرات سازنده خود موجبات بهتر شدن نوشتار حاضر را فراهم کردند، سپاسگزارم.

مراجع

- [1] M. Amintoosi, "Application of Taylor Series in Reducing the Size of Convolutional Neural Networks to Classify Impressionist and Miniature Paintings", *Mathematics and Society*, Vol. 5, No. 1, pp. 1-16, 2020
- [2] S. Sheykhivand, S. Meshgini, and Z. Mousavi, "Automatic Detection of Various Epileptic Seizures from EEG Signal Using Deep Learning Networks," *Computational Intelligence in Electrical Engineering*, Vol. 11, No. 3, pp. 1-12, 2020.
- [3] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, Vol. 2, No. 4, pp. 217-236, 1983.
- [4] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," presented at the Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 2001.
- [5] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-Based Super-Resolution," *IEEE Comput. Graph. Appl.*, Vol. 22, No. 2, pp. 56-65,

- arbitrary neural artistic stylization network," *CoRR*, Vol. abs/1705.06830, / 2017.
- [31] J. Vicente and P. Guillemant, "An image processing technique for automatically detecting forest fire," *International Journal of Thermal Sciences*, Vol. 41, No. 12, pp. 1113 - 1120, 2002.
- [32] P. Barmpoutis, K. Dimitropoulos, K. Kaza, and N. Grammalidis, "Fire Detection from Images Using Faster R-CNN and Multidimensional Texture Analysis," presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.
- [33] P. Li and W. Zhao, "Image fire detection algorithms based on convolutional neural networks," *Case Studies in Thermal Engineering*, Vol. 19, p. 100625, 2020.
- [34] D. Shen, X. Chen, M. Nguyen, and W. Q. Yan, "Flame detection using deep learning," presented at the 4th International Conference on Control, Automation and Robotics (ICCAR), 2018.
- [35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," presented at the Proceedings of the 28th International Conference on Neural Information Processing Systems – Vol. 1, Montreal, Canada, 2015.
- [36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [37] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.
- [38] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: object detection via region-based fully convolutional networks," presented at the Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 2016.
- [39] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, L. B., M. J., S. N., and W. M., Eds., 2016, Springer, pp. 21-37.
- [40] D. Mukhopadhyay, R. Iyer, S. Kadam, and R. Koli, "FPGA Deployable Fire Detection Model for Real-Time Video Surveillance Systems Using Convolutional Neural Networks," in *2019 Global Conference for Advancement in Technology (GCAT)*, 18-20 Oct. 2019 2019, pp. 1-7.
- [41] A. Jadon, M. Omama, A. Varshney, M. S. Ansari, and R. Sharma, "FireNet: A Specialized Lightweight Fire & Smoke Detection Model for Real-Time IoT Applications," ed: arXiv preprint arXiv:1905.11922, 2019.
- presented at the Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2017.
- [20] G. Perarnau, J. Van De Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional gans for image editing," *arXiv preprint arXiv:1611.06355*, 2016.
- [21] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 202x, no. Per-Print, p. 1, 202x.
- [22] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural Style Transfer: A Review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 11, pp. 3365-3385, 2020.
- [23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125-1134.
- [24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242-2251, 2017.
- [25] L. Jiang, C. Zhang, M. Huang, C. Liu, J. Shi, and C. C. Loy, "TSIT: A Simple and Versatile Framework for Image-to-Image Translation," in *16th European Conference on Computer Vision*, Vol. 12348, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm Eds., (Lecture Notes in Computer Science, UK, 2020, pp. 206-222. Springer.
- [26] V. Dumoulin, J. Shlens, and M. Kudlur, "A Learned Representation For Artistic Style, International Conference on Learning Representations 2017 (Conference Track)," 2017.
- [27] M. Amintoosi, "Combining a Regularization Method and the Optimal Brain Damage Method for Reducing a Deep Learning Model Size," *Journal of Machine Vision and Image Processing*, Vol. Accepted, p. (In Persian), 202x. Iranian Society of Machine Vision and Image Processing.
- [28] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, Vol. 6, No. 1, p. 60, 2019/07/06 2019.
- [29] P. Jackson, A. Atapour-Abarghouei, S. Bonner, T. Breckon, and B. Obara, "Style augmentation : data augmentation via style randomization," presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Deep Vision, Long Beach, CA, USA, 2019.
- [30] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens, "Exploring the structure of a real-time,

¹ Neural Style Transfer (NST)

² Gatys

³ Content

⁴ Style

- ^۵ حق نشر تصویر روباه متعلق به مجید مؤمنی مقدم، عضو هیات علمی گروه زیست‌شناسی دانشگاه حکیم سبزواری است.
- ^۶ Object Detection and Localization
- ^۷ Over Fitting
- ^۸ <https://github.com/mamintoosi/MMM-Artistic-photoes>
- ^۹ برای آشنایی بیشتر با معماری این شبکه می‌توان به منابع متعدد موجود منجمله سایت زیر مراجعه کرد:
- VGG16 - Convolutional Network for Classification and Detection (neurohive.io)
- ^{۱۰} Adversarial Generative Networks (GAN)
- ^{۱۱} Ian Goodfellow
- ^{۱۲} Jean Pouget-Abadie
- ^{۱۳} Mehdi Mirza [Momen]: <https://memimo.net>
- ^{۱۴} Generator
- ^{۱۵} Discriminator
- ^{۱۶} Generative Adversarial Networks with Python (machinelearningmastery.com)
- ^{۱۷} Deep Convolutional Generative Adversarial Networks
- ^{۱۸} Reconstruction Error
- ^{۱۹} Pretrained Network
- ^{۲۰} Macro Structures
- ^{۲۱} Micro Structures
- ^{۲۲} Semantic Image Synthesis
- ^{۲۳} Image to Image Translation
- ^{۲۴} Impressionism
- ^{۲۵} Claude Monet
- ^{۲۶} Denormalization
- ^{۲۷} Conditional Instance Normalization
- ^{۲۸} Data Augmentation
- ^{۲۹} Shorten
- ^{۳۰} TensorFlow: <https://www.tensorflow.org>
- ^{۳۱} PyTorch: <https://pytorch.org>
- ^{۳۲} <https://github.com/philipjackson/style-augmentation>
- ^{۳۳} Jakson
- ^{۳۴} Computer Vision and Pattern Recognition Laboratory Homepage (kmu.ac.kr)
- ^{۳۵} Corsican Fire Database (univ-corse.fr)
- ^{۳۶} Computer Vision Based Fire Detection Software (bilkent.edu.tr)
- ^{۳۷} Welcome to the Wildfire Observers and Smoke Recognition Homepage (fesb.hr)
- ^{۳۸} bubblebeam/Inferno-Realtime-Fire-detection-using-CNNs: FPGA Deployable Fire Detection Model for Real-Time Video Surveillance Systems Using Convolutional Neural Networks (github.com)
- ^{۳۹} arpit-jadon/FireNet-LightWeight-Network-for-Fire-Detection (github.com)
- ^{۴۰} tzutalin/labelImg: □□ LabelImg is a graphical image annotation tool and label object bounding boxes in images (github.com)
- ^{۴۱} Bounding Box
- ^{۴۲} Annotating
- ^{۴۳} Ground Truth Annotations
- ^{۴۴} Berkeley DeepDrive
- ^{۴۵} https://www.kaggle.com/solesensei/solesensei_bdd100k
- ^{۴۶} datasets/day2night at main · mamintoosi/datasets (github.com)
- ^{۴۷} OlafenwaMoses/ImageAI: A python library built to empower developers to build applications and systems with self-contained Computer Vision capabilities (github.com)
- ^{۴۸} Scale & Crop
- ^{۴۹} Keras: Keras: the Python deep learning API
- ^{۵۰} Epoch
- ^{۵۱} Validation Set
- ^{۵۲} <https://github.com/mamintoosi/ST-for-DA-in-FD>
- ^{۵۳} Day 2 Night (D2N)

- ⁵⁴ Loss Function
- ⁵⁵ Average Precision (AP)
- ⁵⁶ Intersection over Union
- ⁵⁷ Ground Truth Bounding Boxes
- ⁵⁸ Precision-Recall
- ⁵⁹ True Positive (TP)
- ⁶⁰ False Positive (FP)
- ⁶¹ False Negative (FN)

