

BUET CSE FEST 2023 - DLSPRINT 2.0



1 Mirza Nihal Baig



1 Mahdi Murshed



3 Sowmen Das

¹ Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh
² Statistics, Shahjalal University of Science and Technology, Bangladesh
³ MLMI, University of Cambridge, UK

পাঠকজন ওজনসূচী সীমিত করি, সেটা ও অনুযায়ী সন্তানের দশল প্রকার মন্যমে তার ওজনটি
জ্ঞাতকর রাখা। যে ৩০০টি পিয়ার ছাড়াই যথা ৩০টি প্রকার বস ওজনসূচী পায় নিতের দশল রাখা।
শরীরজনের বিশেষণে মাটি বাপন করে ফেলবারই মন্য প্রকার, বিশেষ করে ক্রিয়াকর্মী তৎপরতা
প্রকার অতিশয় পরিচালনা। ও প্রকারে হারানিকরকার বা খরিজ তার শরীরে শরীরাকর্মী পরিচী চালনা
(ক্ষুধিবর্ধী) উপর দ্রুত ও যথার অংগে তারা পঠিক পড়তা।

ଅବଧା ବିଜ୍ଞାନର ପରବର୍ତ୍ତୀ ପର୍ବରେ ବିଜ୍ଞାନୀମାନଙ୍କୁ ଏହି ବାସନିକ ପ୍ରତିରୋଧ କରାଯେ ଏବଂ ନିମ୍ନଲିଖିତ ସ୍ୱାସ୍ଥ୍ୟ ଅବସ୍ଥା ଧ୍ୟାନ ଦେବାକୁ ପଡ଼ିଥାଏ ।

[illegible][illegible]

পৰিকল্পনা প্ৰকল্পটোৱে ইতিমধ্যে অসমত বহুখণ্ডীয় ব্যৱস্থাৰ অধীনত কাম চালাই আছে। এই ক্ষেত্ৰত ইয়াৰ লক্ষ্য হৈছে দেশৰ ভিতৰলৈ বিনিয়োগ আকৰ্ষণ কৰা আৰু উন্নয়নশীলতা বৃদ্ধি কৰা।

paragraph 84%

অন্য বিভাগের পত্র **paragraph 81%** প্রতিবেশ করছে এবং ১৯
সিদ্ধি হবার অশাধা থাকলে ভাবী ভাবে বিভাগের ওপরই আশ্রয়মান করবে।

[illegible][illegible]

About BaDLAD



- Text-box
- Paragraph
- Image
- Table



Domain	Samples Text-box Paragraph Image Table				
Historical Newspapers	516	25452	38990	1252	67
New Newspapers	96	3978	2507	494	36
Govt Documents	771	44017	2260	762	514
Magazines and Books	18380	123099	162570	7734	594
Lib War Docs	602	7330	3262	55	142
Total	20365	203876	209589	10297	1353

Train Dataset

Domain	Samples Text-box Paragraph Image Table				
Historical Newspapers	345	17611	26571	838	54
New Newspapers	65	3542	1902	237	24
Govt Documents	514	27903	1497	482	301
Magazines and Books	11674	80581	103390	4949	376
Lib War Docs	402	4733	2370	16	104
Total	13000	134370	135730	6522	859

Test Dataset

More Data, Better Accuracy?

	Badlad Train Data	Extended Badlad Data
paragraph	18955	25239
text-box	19502	25764
image	4441	4950
table	892	1039

TABLE I: Frequency table for Previous and Extended Dataset

Previous Train Dataset Size: **20365**

Newly Annotated Data= **6439**

Updated Dataset Size= **20365+ 6439=26804-84=26720**

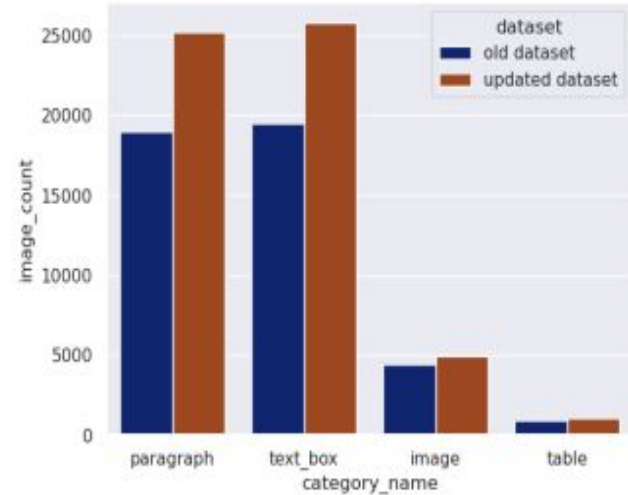


Fig. 1: Frequency Comparison between two versions of Dataset

Is Augmentation Always Better?



➡ CopyPaste+MotionBlur+Rotation

➡ MotionBlur+Rotation

PointRed vs Mask R-CNN

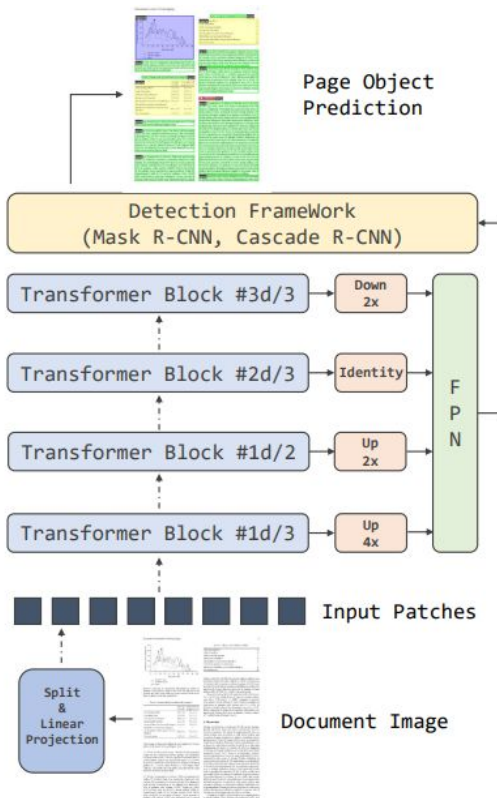
paragraph 100% আমার কোন কষ্ট হয় না।
paragraph 100% আমি আমার কষ্ট কখনোই হয় না, সে প্রমাণ যে দীর্ঘ জীবন করে প্রতি মুহুর্তে সন্তোষ পেয়ে। কষ্ট হয় না—এ কথাটা ঠিক নয়। মানুষের শরীর তো আমার কথা হল, আমার জানো সব কষ্টই সে স্বচ্ছন্দে এবং হাসি মুখে সহ্য নিয়ে পারে, এবং তাই নিয়েছে। শুধু আমার জন্যে কেন, সমস্ত সপোনের জন্যে।
paragraph 100% তাই আমার মা মারা গিয়েছিলেন। সুতরাং বাবার সেবা-ওষুধ, যত্ন-অভিযন্ত্র সব ভারও ছিল তার হাতে। কেউ বলে দেয়নি। স্বৈরাচার এবং সানন্দে সে নিজেই তুলে নিয়েছিল। তার বললে বাবার কাছ থেকেও যে বিশেষ কোন আবার যেন কিবা নিষ্টি ব্যবহার পেয়েছে, তা নয়। বরং যেখানও গেলেন ক্রটি হলে তৎক্ষণাৎ সেটা তাকে কড়াভাবে জানিয়ে দেওয়া হয়েছে।
paragraph 100% পুর দুটোই ছিলাম ভোজনবিলাসী। রোজনকার রান্নার ব্যাপারটা পদ ও পরিমাণ দুটোকেই বেশ ব্যাপক আকারে চালানো হত। কিন্তু 'মাতুল' কামরার মত বিলাসকে আমার কোনদিন প্রভাব মিটনি। চাকর একটা ছিল, বাকী সব অম্পকিনী। আমার একসঙ্গে খেতে বসতাম এবং আশে নিজেই খেতাম। সে অক্লিম দৃষ্টিতে যেত। তারপর কী থাকত, তা নিয়ে আমার কথাটা মাথা ঘামাইনি। মেয়েদের স্বাভাব্য-পরাণ নিয়ে নড়র দেওয়া 'মৌমুখ্যে' পুরুষের লক্ষণ। মাঝে মাঝে খোঁজ নিতে ইচ্ছে হলেও কেমন যেন খোঁজকে বাধত। মনে হত, তাকে করে নিজেই অসুস্থকামি মনিয়ে আন হলে।
paragraph 100% তার ছিল বাবার হাতে। কিন্তু 'ভাল ভাল জিনিসের সোত খতখান, হাত কতটা সফল ছিল না। রান্নার সঙ্গে পুকেটের বিগোল মাছই সেয়ে থাকত, এবং তার সেরে নিজে পড়ত পুরনবুর উপর। বাজারে বেগোবার সময় রোজই একবার ফিফাস করতেন, 'কি কি আছেতে হবে, বৌমা?' প্রকৃতি অমানবিক। 'বৌমা' হো পুঁথিখী না, বৌদনী মায়। কোন কোন জিনিসের প্রয়োজন, ছির কবরার কতী তিনি নিজেই। মন্দাকিনী তা জানত। তাই মনে হত শুধু তানাত, মন্য ভরকারি আপনায় ঘোটা পছন্দ হয়, আনন্দে। মশলাপাতি, বি, সেল সব আছে।
paragraph 100% বি মাসের গোড়াতেই পুরো মাসের মত সংগ্রহ করা হত। সে স্বর্গ তিন নিজে করতেন। মাঝে মাঝে ঘাটতি পড়ত। সেদিনই হত মুশকিল। 'বৌমা' হয়তো ঘিরের টিনটা হাতে করে মাঝে কাঁড়র থেকে বাইরে পা দিয়েছে, শব্দর অমনি গাড়ে উঠতেন, 'ওই তো সেদিন বি নিজে এলাম, এবই মতো শেষ করছে?'
paragraph 99% কিন্তু না, মাথা নিচু করে দাঁড়িয়ে থাকত। কখনো বলত না, ইসলামী পরপূর কদিন রায়ে কটি পরটার বললে লুটি হতোয়ে গলেই ঘিটা যখন খাবার কথা ছিল, তা

(a) Mask-RCNN

paragraph 99% কখন উই-এ আঘাত ২৭/২৮। হঠাৎই শব্দ পড়িয়েছিলম কেবলমতে ভ্রমাত্তে হবার জন্য।
paragraph 100% নির্দেশ ছিল, হঠাৎমতে যদি কোন অবশ্যলী অভিব্যক্ত প্রোমোনেরকে নিয়ন্ত্রণ করতে চান তা হলে তাকে বন্দী অবস্থা হওয়া করবে।
paragraph 100% সাময়িক ২৭ শে মার্চ দুপুর বিরাপিতের মাঝে নিয়ন্ত্রণ করবার জন্য। কিন্তু বাকী তাকে চাপের কতলে ক্যাপ্টেন শিকল নিয়ে বন্ধীকৃত হওয়া করে। পরে বিরাপিতের ভিতর থেকে তুলি করে পাল্লায় ক্যাপ্টেনকে হত্যা করা হয়। ক্যাপ্টেনের সঙ্গীরাও প্রাণ হারায়। সমগ্র ঘণ্টার ক্যাপ্টেনমেন্ট এলাকা আমাতে দেখতে হতো, মনে হতোও ২/৪ খণ্ডীয় বেশী অস্থলন করলে পরভরাম না।
paragraph 100% এটিল ফিকরখান। ঘাই ২৭ই উই-এ দেখবার জন্য। কিন্তু পাকবাহিনী কখন আসের হয়ে পায় বহু করে বেলে এবং সংঘর্ষ হয়। আমি ভারত সীমান্ত ঘুরে ১১ ই এটিল বেলপেলস অসি।
paragraph 100% ২৭ই ২৮ মার্চ বিরসরক্ষার আর ক্যাপ্টেন কে, বি, সিংহ বিরসরক্ষ আইসি, ওনা গার্ডস-এসের সাথে অশান্ত-অশান্তো করে এবং পাক ক্যাপ্টেন দুজনের খতনা বসি। আমি বর্শনা হয়ে চুড়ায়াস অসি এবং মেজর ওসমানকে দুজের কথা বসি। ফিকর গায়া থেকে ওই দিনই ফিরেইয়ে তার ১০ টায় পৌঁছে ঘাই।
paragraph 100% ২৭ই ২৮ মার্চ বাবার থেকে ক্যাপ্টেন শফিউল্লাহর মেসেজ পেলাম যে ওখানে জীবন আটপাড়া এবং ভারী অস্ত্রের অক্রমণ চালাচ্ছে পাক বাহিনী। সে আমার কাছে মতামত চাইল। আমি তাকে কালিঙ্গ চলে আসতে বলি এবং আমাদের বাহিনীকে কালিঙ্গ এবং বাবায়াতের মন্যমানে থাকতে বলি। ওখানেও আমাদের বাহিনী বেশ কিছু অস্ত্র এবং সেনা হারিয়ে ফিরে আসে।
paragraph 100% ২৭ই এটিল পাক বাহিনী কালিঙ্গ লক্ষ্য করে নেয়। ১৪ই এটিল আমি চুড়ায়াস অসি। ১৪ই এটিল পাক বাহিনী ফিরেইয়ে চলে আসে। ১০ ই এটিল আমরা মেহেরপুরে ঘাটি পরিচালনা করে।
paragraph 99% ২৭ই মেহেরপুরে থাকলাম। এটিল হাতে ইয়াবাখী বিরাপিতের। ১৭ই এটিল খুব গোয়ে মুজিবল্লার বৈদ্যনাথকলা ঘাই অস্ত্র-বিহীনদের শব্দ অস্ত্রীনে মোগল দেবার জন্য। ইটিয়ার-এর একটি বোম্পানি নিয়ে গিয়েছিল। বিজ্ঞান ইটিয়ার বাহিনী সব বোকাইয়ে একতর হয়। সমস্ত এটিল মাসেটা সেনা সংগ্রহ করা হয়। যে মাস থেকে কেল্পানী তাল করে নেওয়া হয়।
paragraph 100% ২৭ই ২৮ মার্চ আঘাত পড়িয়েছিল কয়েকজন অভিব্যক্ত কলকালীন পূর্ব পাকিস্তানের সামরিক পরিচালিত পরোয়ানের জন্য রানসনিন এরাপিত জনাব মজবুতমিন আহমেদের বাসায় (ফিনাইসমে) একত্রিক হই ১। বৌতিক এলাকা, এসবিও, মেহেরপুর, ২। আমজাবতিন সিদ্দীকী, এসবিও, মজলী, ৩। ওয়াসিউল ইসলাম, এসবিও, মজলী, ৪। শাহ মোহাম্মদ ফজিল, এসবিও, মেহেরপুর।
paragraph 100% ২৭ই ২৮ মার্চ আঘাত পড়িয়েছিল কয়েকজন অভিব্যক্ত কলকালীন পূর্ব পাকিস্তানের সামরিক পরিচালিত পরোয়ানের জন্য রানসনিন এরাপিত জনাব মজবুতমিন আহমেদের বাসায় (ফিনাইসমে) একত্রিক হই ১। বৌতিক এলাকা, এসবিও, মেহেরপুর, ২। আমজাবতিন সিদ্দীকী, এসবিও, মজলী, ৩। ওয়াসিউল ইসলাম, এসবিও, মজলী, ৪। শাহ মোহাম্মদ ফজিল, এসবিও, মেহেরপুর।
paragraph 100% ২৭ই ২৮ মার্চ আঘাত পড়িয়েছিল কয়েকজন অভিব্যক্ত কলকালীন পূর্ব পাকিস্তানের সামরিক পরিচালিত পরোয়ানের জন্য রানসনিন এরাপিত জনাব মজবুতমিন আহমেদের বাসায় (ফিনাইসমে) একত্রিক হই ১। বৌতিক এলাকা, এসবিও, মেহেরপুর, ২। আমজাবতিন সিদ্দীকী, এসবিও, মজলী, ৩। ওয়াসিউল ইসলাম, এসবিও, মজলী, ৪। শাহ মোহাম্মদ ফজিল, এসবিও, মেহেরপুর।

(b) PointRed

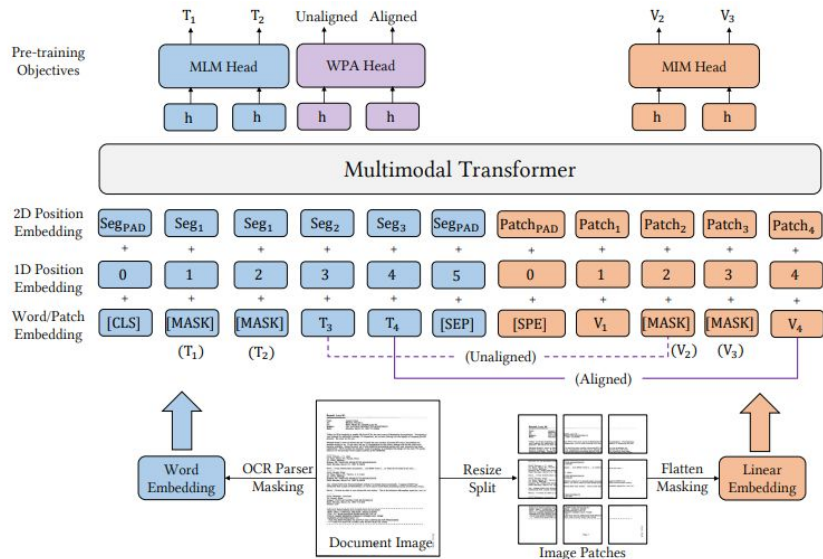
DiT: Self-supervised Pre-training for Document Image Transformer



Hyperparameters	Values
Batch Size	2
Nominal Batch Size	40
Iterations	50000
Solver Gamma	0.001

Hyperparameters used for DiT

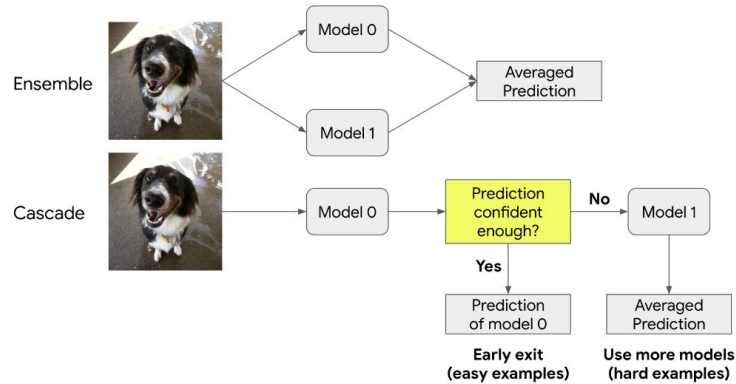
LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking



Hyperparameters	Values
Batch size	3
Nominal Batch Size	27
Solver Gamma	0.0001
Gradient Accumulation Steps	1
Number of iterations	92000

TABLE II: Hyperparameters used for training LayoutLMv3

Ensemble Technique (NMS)



Model Name	Accuracy (Dice Score)	Inference Time
DiT (50k iteration)	.89349	1 hr 54 mins
Layoutlmv3(50k iteration)	.89554	2 hr 4 mins
Ensemble (DiT+Layoutlmv3)	.89677	14 hr 17 mins

Results

After 5k iteration

Model	Submission Score
Mask R-CNN 50 FPN	0.85574
Mask R-CNN 101 FPN	0.86017
PointRend	0.73567
DiT	0.86943
LayoutLMv3	0.8769

TABLE III: Public Leaderboard scores for different models

Thank You

Feel free to ask any questions!

**GitHub repo will be updated soon*