

Learning to Optimize Radiotherapy Plans via Fluence Maps Diffusion Model Generation and LSTM-based Optimization

Isabella Poles^{1,2*}, Simon Arberet², Riqiang Gao², Martin Kraus³, Marco D. Santambrogio¹, Florin C. Ghesu³, Ali Kamen², and Dorin Comaniciu²

¹ Politecnico di Milano, Milan, Italy
isabella.poles@polimi.it

² Digital Technology and Innovation, Siemens Healthineers, Princeton, NJ, USA

³ Digital Technology and Innovation, Siemens Healthineers, Erlangen, Germany

Abstract. Volumetric Modulated Arc Therapy (VMAT) is a cornerstone of modern radiation therapy, enabling highly conformal tumor irradiation and healthy-tissue sparing. Yet, its planning solves inverse and nested optimization for multi-leaf collimators, monitor units and dose parameters, while enforcing their consistency to ensure mechanical deliverability. Nevertheless, this process often requires repeated re-optimization when treatment configurations change, resulting in substantial planning time per patient. To address these problems, we present a diffusion-driven Learning-to-Optimize (L2O) method for end-to-end VMAT planning. A distribution-matching distilled diffusion model learns a clinically feasible manifold of fluence maps, enabling their one-shot generation. On top of this, an LSTM-based L2O module learns gradient update dynamics to swiftly refine fluence maps toward prescribed dose objectives during inference. Experimental results on clinical and public prostate cancer cohorts demonstrate improved planning efficiency, flexibility, and machine deliverability over currently available end-to-end VMAT planners.

Keywords: VMAT Planning · Diffusion Models · Learning-to-Optimize.

1 Introduction

Radiation Therapy (RT) is a core component of modern cancer treatment, enabling irradiation of tumors to impair cell proliferation and halt disease progression [13]. Intensity Modulated RT (IMRT) and Volumetric Modulated Arc Therapy (VMAT) are the most widely adopted techniques [21]. While both are effective, VMAT enhances target dose conformity and healthy-tissue sparing via fast, continuous radiation modulation as the machine’s Multi-Leaf Collimator (MLC) rotates around the patient. Despite this, its planning is a high-dimensional inverse problem with nested optimization of the MLC radiation aperture, Monitor Units (MU) intensity, and dose objectives, which are adjusted until a clinically

* Corresponding author

acceptable plan is achieved. Moreover, machine-dose *consistency* must be enforced to ensure that the planned dose remains deliverable under machine constraints, further complicating the process. Consequently, current pipelines often require tens of minutes per patient and per treatment-configuration change [3].

Motivated by the need for efficient, flexible, and dose-machine-consistent solutions, we shift the VMAT planning paradigm from a conventional optimization to an end-to-end diffusion-based Learning-to-Optimize (L2O) formulation.

Some recent end-to-end VMAT planners optimize MLC apertures and MU via gradient-based Direct Aperture Optimization (DAO) [25] and (un)constrained optimizers [4, 5, 18, 20, 32], allowing DL-based dose predictors [6, 7, 9] and engine simulation [19] to be embedded in differentiable pipelines, potentially accelerating planning. Other approaches cast planning as a decision-making problem using Reinforcement Learning (RL) [12, 15], or as a supervised inference task enabling single forward-pass plan prediction [1, 30]. Nevertheless, the literature highlights three key limitations. First, certain DAO planners report piecewise-differentiable and plateau-heavy loss landscapes due to MLC quantization and gap constraints, leading to illusory/slow convergence, which hinders practical deployment [29]. Second, some RL-based approaches adjust machine parameters through discrete or partially continuous actions that inadequately explore the full MLC search space [12]. Convolutional and multi-agent policies solve the issue, enabling continuous dose-rate and MLC modulation, but requiring large RT-plan datasets for generalization [23] and remaining restricted to leaf-sequencing tasks based on pre-optimized fluence maps beam-intensity patterns [8]. Third, supervised approaches can be hindered by the non-uniqueness of machine-parameter-dose mappings and limited adaptability to changing planning conditions [11].

To address these gaps, we propose a unified fluence-map generation L2O method achieving fast, flexible, and clinically feasible end-to-end VMAT planning, as shown in Figure 1. Indeed, fluence maps provide a differentiable parameterization of the beam-intensity modulation achievable through feasible MLC motion and MU patterns, yielding a smooth representation that supports stable gradient-based updates while promoting physically deliverable patterns [2]. Therefore, our primary goal is to update fluence maps iteratively, leveraging the *smooth* gradient landscape computed between a DL-predicted and clinically desired dose to satisfy dose-machine-consistent clinical objectives and to *learn* efficient *optimization* dynamics. To achieve this, we employ diffusion models to model the non-unique distribution of clinically feasible planning solutions, and leverage **distribution-matching distillation for efficient single-shot fluence map generation**. Operating in this manifold without relying on target fluence map initializations, the LSTM-based **L2O module** learns the history of the optimization gradients during training, **finding the fast and saddle-points free pathway to plan solution** during inference. Unlike existing methods, **our approach is initialization agnostic, efficient, flexible** to changing objectives **without retraining** and **guarantees plan delivery**.

We summarize our contributions as follows: (1) FMD: a one-shot Fluence Map Diffusion model trained via distribution-matching distillation to generate

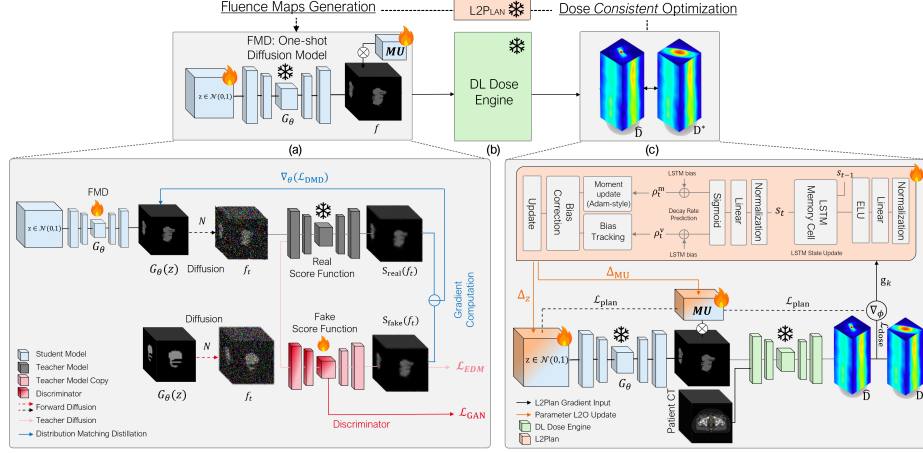


Fig. 1. Overview of our method from the FMD model, which generates fluence maps in one-shot (a), to the L2PLAN optimizer (c), which learns to optimize fluence maps so that the corresponding predicted dose (b) matches the target plan.

VMAT fluence maps and a manifold of non-unique plan solutions; (2) L2PLAN: a novel LSTM-based L2O VMAT Plans module that learns iterative update dynamics to efficiently refine fluence maps toward prescribed dose objectives without tuned initializations or retraining; (3) A validation on three prostate cancer clinical and public patient cohorts demonstrating improved efficiency, flexibility to changing dose objectives, and plan deliverability verified via machine leaf sequencing compared to existing optimization and learning-based approaches.

2 Methodology

Our methodology combines diffusion-based fluence generation with L2O refinement to enforce dose-machine consistency, ensuring a prescribed dose $\mathbf{D}^*$ admits at least one feasible fluence map $\mathbf{f} \in \mathcal{F}_{\text{del}}$, where $\mathcal{F}_{\text{del}}$ denotes the set of machine-constrained fluence maps defined by limits on leaf motion, dose rate, and gantry rotation, and validated via MLC Leaf Sequencing (LS).

First, we distill a diffusion model trained on $\mathcal{F}_{\text{del}}$ into a one-shot generator G_θ that rapidly samples a plausible initial fluence $\mathbf{f} = G_\theta(\mathbf{z})$ from latent noise $\mathbf{z} \sim \mathcal{N}(0, I)$, without iterative denoising. Second, we train an LSTM-based optimizer to update $\mathbf{z}$ and a $\mu\mathbf{U}$ parameter to scale $\mathbf{f}$ intensity, $\phi = \{\mathbf{z}, \mu\mathbf{U}\}$, using the dose gradients history between the planned dose $\mathbf{D}^*$ and the predicted one $\hat{\mathbf{D}}$, obtained from $\mathbf{f}$ and a patient Computed Tomography (CT) image via a differentiable dose engine [19]. The diffusion and dose modules are trained separately and kept frozen during LSTM training; all modules remain frozen at inference, enabling rapid refinement of $\mathbf{f}$ via ϕ while preserving deliverability. An overview is provided in Figure 1.

2.1 FMD: Fluence Maps Diffusion Model Generation in One-shot

Despite their ability to model high-quality deliverable VMAT fluence maps, diffusion models require expensive iterative denoising, making distillation into a one-shot generator imperative for efficient planning. Therefore, to guarantee high-quality generation and facilitate the subsequent distillation process, we first train a Dhariwal-style UNet as the teacher diffusion model, employing four down-sampling stages with residual blocks, group normalization, SiLU activations, and a symmetric skip-connected decoder following EDM [16]. In particular, in the forward diffusion process, a fluence map $\mathbf{f} \sim \mathcal{F}_{\text{real}}$ is progressively corrupted by injecting Gaussian noise across T continuous noise levels, such that at noise level t it follows the marginal distribution $\mathcal{F}_{\text{real},t}(\mathbf{f}_t) = \int \mathcal{F}_{\text{real}}(\mathbf{f}) z_t(\mathbf{f}_t | \mathbf{f}) d\mathbf{f}$, where $z_t(\mathbf{f}_t | \mathbf{f}) = \mathcal{N}(\alpha_t \mathbf{f}, \sigma_t^2 \mathbf{I})$ and $\alpha_t, \sigma_t > 0$ are defined by the EDM noise schedule. Teacher reverse diffusion is learned via denoising score matching by minimizing:

$$\mathcal{L}_{\text{EDM}} = \mathbb{E}_{\mathbf{f} \sim \mathcal{F}_{\text{real}}, t \sim [0,T]} \|\mu_{\text{real}}(\mathbf{f}_t, t) - \mathbf{f}\|_2^2, \quad (1)$$

which estimates a clean fluence map μ_{real} from noisy ones and, implicitly, the score function $\mathbf{s}_{\text{real}}(\mathbf{f}_t, t) = \nabla_{\mathbf{f}_t} \log \mathcal{F}_{\text{real},t}(\mathbf{f}_t)$ of the diffused distribution $\mathcal{F}_{\text{real}}$.

To enable one-shot sampling, we distill the teacher into a single-step generator G_θ via Distribution Matching Distillation (DMD) [31], minimizing the expected Kullback-Leibler (KL) divergence between the teacher $\mathcal{F}_{\text{real},t}$ and student $\mathcal{F}_{\text{fake},t}$ diffused data distributions. Since G_θ is trained by minimizing KL loss between the teacher-student diffusion distributions, whose likelihood is intractable, we directly compute the gradient of the KL loss, which depends just on the difference between their score functions, making explicit density evaluation unnecessary:

$$\nabla_\theta \mathcal{L}_{\text{DMD}} = -\mathbb{E}_{\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}), t} \left[(\mathbf{s}_{\text{real}}(\mathbf{f}_t, t) - \mathbf{s}_{\text{fake}}(\mathbf{f}_t, t)) \frac{\partial G_\theta(\mathbf{z})}{\partial \theta} \right], \quad (2)$$

where $\mathbf{f}_t = N(G_\theta(\mathbf{z}), t)$, N denotes the forward noising process, and $\mathbf{s}_{\text{fake}}$ is estimated by a teacher copy trained online on the G_θ generated samples. Since the G_θ never observes real data, inheriting approximation errors from the teacher, we anchor G_θ to real data, by training a discriminator $\mathcal{D}_{\text{GAN}}$ on noisy samples [31]:

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{\mathbf{f} \sim \mathcal{F}_{\text{real}}, t} [\log \mathcal{D}_{\text{GAN}}(N(\mathbf{f}_t, t))] + \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}), t} [\log(1 - \mathcal{D}_{\text{GAN}}(N(G_\theta(\mathbf{z}), t)))] \quad (3)$$

which gradient jointly with $\nabla_\theta \mathcal{L}_{\text{DMD}}$ update G_θ . The resulting updates are independent of the teacher’s sampling trajectories and mitigate score approximation errors, allowing G_θ to match the teacher’s sample quality in a single step.

2.2 L2PLAN: Learning-to-Optimize Dose Consistent RT Plans

After generating an initial $\mathbf{f}$, we enforce consistency by $\phi = \{\mathbf{z}, \mathbf{MU}\}$ updating via backpropagation of the plan loss $\mathcal{L}_{\text{plan}}(\hat{\mathbf{D}}, \mathbf{D}^*)$, driving the predicted dose $\hat{\mathbf{D}}$ toward the target one $\mathbf{D}^*$. Although ϕ end-to-end optimization is feasible, standard gradient-based methods suffer from instability/saddle-point stagnation, as

even simple dose objectives induce a non-convex landscape when composed with the nonlinear fluence-to-dose mapping, which static hyperparameters struggle to adapt to. As a result, optimization may drift toward numerically optimal yet physically inconsistent solutions, characterized by needle-like apertures or spiky MU profiles that violate machine constraints.

To address this issue, we adopt an L2O strategy that jointly trains a recurrent optimizer to learn task-adaptive update rules that escape saddle points and enable fast dose-driven parameter refinement. L2PLAN consists of an *inner optimizee loop*, which iteratively updates ϕ through learned hyperparameterized update rules, and an *outer meta-optimization loop*, which learns the parameters of the meta-optimizer itself. Starting from randomly initialized meta-optimizer parameters at iteration $k = 0$, the outer loop performs K_0 meta-optimization steps, where each outer update is driven by the outcome of an inner optimizee unrolled for $k_i = 10$ steps. In the inner loop, at step k , the dose objective gradient $\mathbf{g}_k = \nabla_{\phi_k} \mathcal{L}_{\text{plan}}$ is normalized, passed through a learnable linear preprocessing function $\delta(\cdot)$ and used to update a latent optimizer state $\mathbf{s}_k$ which implicitly track the gradient history via an LSTM with two recurrent layers:

$$\mathbf{s}_k = \text{LSTM}(\delta(\mathbf{g}_k), \mathbf{s}_{k-1}). \quad (4)$$

Conditioned on the current state $\mathbf{s}_k$, the meta-optimizer predicts hyperparameters inspired by adaptive first-order optimization methods, namely the first and second order momentum decay rates β_k and γ_k , through meta-learnable linear projections $l(\cdot)$ followed by σ sigmoid activations to ensure valid ranges:

$$\beta_k = \sigma(l_\beta(\mathbf{s}_k)), \quad \gamma_k = \sigma(l_\gamma(\mathbf{s}_k)). \quad (5)$$

These parameters are initialized to standard Adam values at $k = 0$ and subsequently adapted according to the gradient signals produced during the learned optimization process. Using the β_k , γ_k and $\mathbf{g}_k$, moment estimates are updated as $\mathbf{m}_k = \beta_k \mathbf{m}_{k-1} + (1 - \beta_k) \mathbf{g}_k$ and $\mathbf{v}_k = \gamma_k \mathbf{v}_{k-1} + (1 - \gamma_k) \mathbf{g}_k^2$, with bias correction applied. Given the predicted update direction, the resulting optimizee step $\Delta\phi_k$ and parameter update ϕ_{k+1} are computed as:

$$\Delta\phi_k = -\hat{\mathbf{m}}_k / (\sqrt{\hat{\mathbf{v}}_k} + \epsilon) \rightarrow \phi_{k+1} = \phi_k + \Delta\phi_k. \quad (6)$$

After k_i inner-loop updates are performed, $\mathcal{L}_{\text{plan}}$ updates the outer LSTM parameters. The trained meta-optimizer is then frozen, guaranteeing the $\mathbf{D}^*$ clinical objective with the fewest k prediction steps $\Delta\phi_k$ during planning.

2.3 Overall Framework

In our approach, we formulate VMAT planning as a Maximum A Posteriori (MAP) problem which estimates $\phi = \{\mathbf{z}, \mathbf{MU}\}$ values most consistent with the observed $\mathbf{D}^*$ while respecting some prior assumptions. The MAP objective $\mathcal{L}_{\text{plan}} = \mathcal{L}_{\text{dose}} + \lambda_{\mathbf{z}} \mathcal{L}_{\text{cont}}^{\mathbf{z}} + \lambda_{\mathbf{MU}} \mathcal{L}_{\text{cont}}^{\mathbf{MU}}$ is minimized with the learned meta-optimizer over 100 steps k , consistently achieving convergence. Here, the likelihood term

Table 1. Results on the REQUITE dataset. “Optimizer” refers to the strategy enforcing dose-machine consistency; “Generator” denotes the fluence-map model; “Steps” the number of optimization iterations; and “LS” rule-based Leaf Sequencing results. The best results are in bold, while (*) indicates 0.001 p-value significant difference.

Optimizer	Generator	Steps	FID ↓	MSE ↓	SSIM ↑	PSNR ↑	MAE ↓	MAE _{PTV} ↓	MAE _{OARs} ↓	Time ↓
<i>Standard VMAT Optimizers</i>										
RMSProp [4]	FMD	300	17.52	-	0.940±0.040*	32.66±4.35*	0.173±0.065*	1.151±0.224*	0.422±0.298*	504.0±1.1*
L-BFGS [28]	FMD	300	17.52	-	0.971±0.008*	37.91±1.67*	0.099±0.039*	0.638±0.198*	0.255±0.087*	890.4±3.5*
SGD _M [4]	FMD	300	17.52	-	0.978±0.009*	38.48±1.84*	0.089±0.024*	0.431±0.292*	0.183±0.092*	508.2±1.6*
Adam [18]	FMD	300	17.52	-	0.985±0.007*	39.47±2.56*	0.085±0.032*	0.720±0.270*	0.231±0.131*	505.1±1.2*
DAO [25]	-	2000	-	-	0.919±0.071*	28.02±4.28*	0.324±0.070*	1.061±0.545*	0.747±0.071*	2966.4±9.4*
Adam [18]	FMD	2000	17.52	-	0.998±0.001	50.90±3.24	0.019±0.009	0.170±0.070	0.060±0.041	2339.2±7.2
<i>Generative Models</i>										
Adam [18]	VQ-VAE [10]	300	-	5e-3	0.988±0.004*	37.71±1.49*	0.080±0.028*	1.221±0.329*	0.395±0.194*	547.6±1.5*
Adam [18]	StyleGAN2 [17]	300	32.94	-	0.979±0.006*	38.40±1.51*	0.113±0.044*	0.920±0.190*	0.440±0.280*	469.0±2.0*
Adam [18]	D2O [33]	300	23.76	-	0.963±0.125*	38.06±6.46*	0.098±0.111*	0.963±0.397*	0.511±0.623*	487.2±1.9*
Adam [18]	FMD	300	17.52	-	0.985±0.007*	39.47±2.56*	0.085±0.032*	0.720±0.270*	0.231±0.131*	505.1±1.2*
<i>L2O Optimizers</i>										
VeLO [14]	FMD	300	17.52	-	0.985±0.015*	40.62±3.75*	0.066±0.018*	0.471±0.033*	0.243±0.120*	461.1±3.1*
μ LO [26]	FMD	300	17.52	-	0.972±0.030*	40.54±0.95*	0.091±0.021*	0.222±0.040*	0.221±0.193*	458.3±1.5*
CoordMath [22]	FMD	300	17.52	-	0.981±0.022*	39.66±1.09*	0.097±0.017*	0.345±0.048*	0.299±0.201*	467.1±1.3*
HyperAdam [27]	FMD	300	17.52	-	0.990±0.020*	41.72±1.11*	0.054±0.011*	0.207±0.090*	0.255±0.107*	450.3±1.0*
L2PLAN	FMD	300	17.52	-	0.998±0.001	50.61±1.84	0.026±0.009	0.149±0.076	0.061±0.039	478.0±0.9
L2PLAN	FMD	100	17.52	-	0.993±0.003	45.33±1.62	0.059±0.019	0.445±0.008	0.154±0.087	159.9±0.5
L2PLAN_{LS}	FMD	100	17.52	-	0.993±0.002	45.22±1.53	0.058±0.019	0.445±0.098	0.163±0.084	161.0±0.4

$\mathcal{L}_{\text{dose}} = \|\hat{\mathbf{D}} - \mathbf{D}^*\|_1$ enforces fidelity to the target dose, while the continuity priors $\mathcal{L}_{\text{cont}}^{\mathbf{z}} = \sum_{\text{CP}} \|\mathbf{z}_{\text{CP}+1} - \mathbf{z}_{\text{CP}}\|_1$ and $\mathcal{L}_{\text{cont}}^{\text{MU}} = \sum_{\text{CP}} \|\mathbf{MU}_{\text{CP}+1} - \mathbf{MU}_{\text{CP}}\|_1$ encourage smooth transitions across Control Points (CPs). We use fixed regularization weights, $\lambda_{\mathbf{z}} = 5$ and $\lambda_{\text{MU}} = 1$, without patient-specific tuning.

3 Experiments

3.1 Experimental Setup

Dataset. We curated 12,469 single-arc prostate VMAT plans from the public REQUITE dataset [24], covering whole-prostate, seminal-vesicle, post-surgical prostate-bed, and pelvic lymph-node PTVs, bladder and rectum OARs segmented using AI-Rad Companion and an in-house C++ pipeline. Across the cohort, plans have fixed 30° collimator and 0° couch angles. Each clockwise arc spans nearly a full rotation and is discretized into 178 CPs. All treatments were planned in Eclipse using the RapidPlan prostate SBRT model and the Varian Millennium 120 (M120) MLC with 60 leaf pairs, 5mm-10mm wide, in the central 40 and the outer 20 pairs, respectively.

Implementation Details. Treatment plans undergo data augmentation by generating synthetic single-arc plan variants using Eclipse TPS with randomized field size, gantry angle variations, and 0.5×-1.5× fluence-map intensity scaling relative to the original value. Diffusion-model and LSTM-based meta-optimizations are performed using Adam (learning rate $1 \cdot 10^{-5}$, momentum 0.9) with batch sizes of 256 and 4, respectively, implemented in PyTorch (1.13.0) on 4×NVIDIA H100 GPUs (80GB). The dataset is split per patient into 12,153-79-79 train-validation-test cases, and results are reported with paired *t*-test analysis.

Table 2. Effectiveness of each module in our method. The best results are in bold, while the (*) and (†) indicate 0.001 and 0.05 p-value significant differences.

FMD Loss			Metric	
$\mathcal{L}_{EDM}$	$\mathcal{L}_{DMD}$	$\mathcal{L}_{GAN}$	Steps↓	FID↓
✓	×	×	18	15.40
✓	✓	×	1	28.18
✓	×	✓	1	22.76
✓	✓	✓	1	17.52

L2PLAN Loss			Metric		
$\mathcal{L}_D$	$\mathcal{L}_{cont}^z$	$\mathcal{L}_{cont}^{MU}$	PSNR ↑	MAE _{PTV} ↓	MAE _{OARs} ↓
✓	×	×	43.22±1.51*	0.509±0.197*	0.168±0.086*
✓	✓	×	44.76±1.47†	0.469±0.198†	0.163±0.086†
✓	×	✓	44.56±1.01†	0.480±0.178†	0.162±0.016†
✓	✓	✓	45.33±1.62	0.445±0.108	0.154±0.087

3.2 Comparison with Currently Available VMAT Planners

We compared our method against RT, latent-variable generative, and L2O baselines. As shown in Table 1, L2PLAN achieves higher accuracy with fewer optimization iterations than standard approaches. Specifically, it improves PTV coverage in Mean Absolute Error (MAE) (0.720 *v.s.* 0.149 MAE_{PTV}) and OARs sparing (0.231 *v.s.* 0.061 MAE_{OARs}) over unconstrained RT planners with the same 300 iterations budget, while reaching comparable performances but with 7× less steps. Notably, Peak Signal-to-Noise Ratio (PSNR) improves from 28.02 to 45.33 and MAE from 0.324 to 0.059 in just 159s and 20× fewer steps than the DAO ones. Importantly, unlike RL-based planners dependent on large RT datasets, L2PLAN is trained with only 180 outer-loop meta-optimization steps. This underscores effective escape from saddle points and faster convergence in training and inference. Furthermore, it is noteworthy that the generative fluence map manifold fidelity influences the accuracy of the dose matching. Indeed, despite the inherent differences between Frechet Inception Distance (FID) and Mean Squared Error (MSE) evaluation, our FMD one-shot generator achieves superior dose consistency than VAE-, GAN- and one-shot diffusion-based alternatives surpassing the leading diffusion model solution both in FID, Structural Similarity Index Measure (SSIM) by 2.2% (0.963 *v.s.* 0.985) and in PSNR by 1.41 (38.06 *v.s.* 39.47) at 300 steps Adam convergence. Remarkably, while alternative optimizers are constrained by stable initialization of optimization parameters, L2PLAN operates without it. Nevertheless, our method also outperforms the most recent L2O strategies by 2.1% in SSIM (0.993 *v.s.* 0.972), 4.71 in PSNR (45.33 *v.s.* 40.54) and 3.2% in MAE (0.059 *v.s.* 0.091) being statistically different.

3.3 Ablation Study

Table 2 reports ablations of the FMD and L2PLAN losses. Using EDM as a teacher, reducing denoising steps slightly increases FID. However, $\mathcal{L}_{DMD}$ and $\mathcal{L}_{GAN}$ act complementarily: the former preserves alignment with the teacher, while the latter corresponds alone to the D2O model [33], enhances perceptual realism. Their combination achieves the best trade-off in manifold fidelity (17.52 *v.s.* 15.40 FID) and efficiency (1 *v.s.* 18 steps). On the optimization side, introducing continuity priors to $\mathcal{L}_D$ progressively improves planning quality. $\mathcal{L}_{cont}^z$ increases PSNR by 1.54 and reduces MAE_{PTV}, while $\mathcal{L}_{cont}^{MU}$ further stabilizes OARs errors. Combining both priors yields the best overall performance.

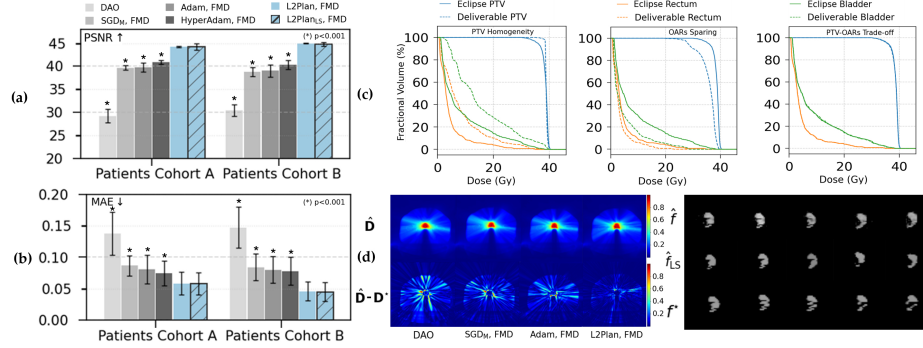


Fig. 2. Results on two private patients cohorts (a, b), flexibility DVH analysis after LS (c), visual results of L2PLAN dose ($\hat{\mathbf{D}}$), the comparison with its target ($\mathbf{D}^*$) and a set of contiguous CPs of L2PLAN-, LS- and target fluence maps ($\hat{f}$, $\hat{f}_{LS}$, f^*)(d).

3.4 Results with Other Datasets

Since L2PLAN acts as an optimizer, we expect its learned update dynamics to transfer to new prostate cohorts and objectives without retraining, provided the task losses remain similar. We evaluate L2PLAN on two private prostate cohorts from different institutions (135 and 13 patients), preprocessed as REQUITE but planned with different objectives (Figure 2 (a,b)). Across both cohorts, our approach improves PSNR and MAE by up to 4.01 and 0.7, respectively, over standard RT planners. It also surpasses the best L2O baseline by 3.41 and 4.65 PSNR and 0.016 and 0.032 MAE on cohorts A-B, demonstrating generalizability.

3.5 Results on Clinically Oriented Application Tasks

We evaluated L2PLAN for plan consistency and flexibility in RT practice translation. Using a rule-based LS tool, we confirm that our optimized fluence maps satisfy Varian M120 MLC constraints, with statistically significant pre-/post-LS agreement across REQUITE (Table 1) and private cohorts (Figure 2 (a,b,d)). For flexibility, we add a $\mathcal{L}_1$ term to $\mathcal{L}_{\text{plan}}$ to promote 38Gy PTV homogeneity or OAR sparing. Figure 2 (c) shows adaptation without retraining: enforcing PTV homogeneity yields 0.041 HI, 38.23 D₉₅, and 39.52 D₂, with reduced OAR sparing (rectum/bladder 8.23/13.04 D_{mean}), while prioritizing OAR protection lowers rectum/bladder D_{mean} to 3.60/ 4.52, with HI increasing to 0.233. Focusing on clinical application, these results show how L2PLAN can adapt consistently and on-the-fly to various objectives, directly supporting RT physicians in practice.

4 Conclusion

We present a diffusion L2O method for end-to-end VMAT planning. A distilled diffusion model learns a clinically feasible fluence manifold, while an LSTM-based optimizer captures adaptive gradient dynamics satisfying dose objectives

efficiently. Unlike most recent planners that suffer from plateau-dominated landscapes, rely on tuned initialization, and compute-intensive re-optimization when objectives change, our validation on public and private datasets demonstrates initialization-agnostic, saddle-point-free, flexible, and machine-consistent planning, bringing VMAT planning one step closer to more practical deployment. Future work will extend to more complex sites and multi-arc plans, and one-shot latent diffusion will reduce optimized parameters and improve efficiency.

Disclaimer. The concepts and information in this paper are based on research results that are not commercially available. Future availability cannot be guaranteed.

Acknowledgments. All contributors to the REQUITE project are acknowledged, including the patients, clinicians, and nurses. The consortium consists of D. Azria, E. Briers, J. Chang-Claude, A. M. Dunning, R. M. Elliott, C. Faivre-Finn, S. Gutierrez-Enriquez, K. Johnson, Z. Lingard, T. Rancati, T. Rattay, B. S. Rosenstein, D. De Ruysscher, P. Seibold, E. Sperk, R. P. Symonds, H. Stobart, C. Talbot, A. Vega, L. Veldeman, T. Ward, A. Webb and C. M. L. West.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Arberet, S., Gao, R., Kraus, M., Ghesu, F.C., Verbakel, W., Diallo, M., Magliari, A., Karuppusamy, V., Beriwal, S., Kamen, A., et al.: Ai end-to-end radiation treatment planning under one second. arXiv preprint arXiv:2603.06338 (2026)
2. Arberet, S., Ghesu, F.C., Gao, R., Kraus, M., Sackett, J., Kuusela, E., Kamen, A.: Beam’s eye view to fluence maps 3d network for ultra fast vmat radiotherapy planning. *Medical Physics* **52**(5), 3183–3190 (2025)
3. Claessens, M., Oria, C.S., Brouwer, C.L., Ziemer, B.P., Scholey, J.E., Lin, H., Witztum, A., Morin, O., El Naqa, I., Van Elmpt, W., et al.: Quality assurance for ai-based applications in radiation therapy. In: *Seminars in radiation oncology*. vol. 32, pp. 421–431. Elsevier (2022)
4. Dubois, P.: Radiotherapy dosimetry: A review on open-source optimizer. arXiv preprint arXiv:2305.18014 (2023)
5. Dursun, P., Hong, L., Jhanwar, G., Huang, Q., Zhou, Y., Yang, J., Pham, H., Cervino, L., Moran, J.M., Deasy, J.O., et al.: Automated vmat treatment planning using sequential convex programming: algorithm development and clinical implementation. *Physics in Medicine & Biology* **68**(15), 155006 (2023)
6. Feng, Z., Wen, L., Cui, J., Wu, X., Xiao, J., Peng, X., Shen, D., Wang, Y.: Leveraging visual prompt with diffusion adversarial network for radiotherapy dose prediction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 302–312. Springer (2025)
7. Gao, R., Arberet, S., Kraus, M., Liu, H., Verbakel, W.F., Comaniciu, D., Ghesu, F.C., Kamen, A.: Generative ai helps radiotherapy planning with user preference. arXiv preprint arXiv:2512.08996 (2025)
8. Gao, R., Ghesu, F.C., Arberet, S., Basiri, S., Kuusela, E., Kraus, M., Comaniciu, D., Kamen, A.: Multi-agent reinforcement learning meets leaf sequencing in radiotherapy. In: *International Conference on Machine Learning*. pp. 14723–14746. PMLR (2024)

9. Gao, R., Lou, B., Xu, Z., Comaniciu, D., Kamen, A.: Flexible-cm gan: towards precise 3d dose prediction in radiotherapy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 715–725 (2023)
10. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., et al.: Maisi: Medical ai for synthetic imaging. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4430–4441. IEEE (2025)
11. Heilemann, G., Zimmermann, L., Nyholm, T., Simkó, A., Widder, J., Goldner, G., Georg, D., Kuess, P.: Ultra-fast, one-click radiotherapy treatment planning outside a treatment planning system. *Physics and Imaging in Radiation Oncology* **33**, 100724 (2025)
12. Hrinivich, W.T., Lee, J.: Artificial intelligence-based radiotherapy machine parameter optimization using reinforcement learning. *Medical physics* **47**(12), 6140–6150 (2020)
13. Jaffray, D.A., Knaul, F., Baumann, M., Gospodarowicz, M.: Harnessing progress in radiotherapy for global cancer control. *Nature cancer* **4**(9), 1228–1238 (2023)
14. Janson, P., Thérien, B., Anthony, Q., Huang, X., Moudgil, A., Belilovsky, E.: Pylo: Towards accessible learned optimizers in pytorch. arXiv preprint arXiv:2506.10315 (2025)
15. Kafeei, P., Cappart, Q., Renaud, M.A., Chapados, N., Rousseau, L.M.: Graph neural networks and deep reinforcement learning for simultaneous beam orientation and trajectory optimization of cyberknife. *Physics in Medicine & Biology* **66**(21), 215002 (2021)
16. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems* **35**, 26565–26577 (2022)
17. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8110–8119 (2020)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Kraus, M.F., Gao, R., Arberet, S., Ghesu, F.C., Kamen, A., Comaniciu, D.: Single shot full plan deep learning dose computation for radiation therapy using spherical harmonics. *Medical Physics* **53**(1), e70223 (2026)
20. Liu, C., Ni, X., Jin, X., Si, W.: Neuraldao: Incorporating neural network generated dose into direct aperture optimization for end-to-end imrt planning. *Medical physics* **48**(10), 5624–5638 (2021)
21. Liu, H., Sintay, B., Pearman, K., Shang, Q., Hayes, L., Maurer, J., Vanderstraeten, C., Wiant, D.: Comparison of the progressive resolution optimizer and photon optimizer in vmat optimization for stereotactic treatments. *Journal of applied clinical medical physics* **19**(4), 155–162 (2018)
22. Liu, J., Chen, X., Wang, Z., Yin, W., Cai, H.: Towards constituting mathematical structures for learning to optimize. In: International conference on machine learning. pp. 21426–21449. PMLR (2023)
23. Mekki, L., Hrinivich, W.T., Lee, J.: Dual-arc vmat machine parameter optimization for localized prostate cancer using deep reinforcement learning. *Physics in Medicine & Biology* **70**(22), 225007 (2025)
24. Seibold, P., Webb, A., Aguado-Barrera, M.E., Azria, D., Bourgier, C., Brengues, M., Briers, E., Bultijck, R., Calvo-Crespo, P., Carballo, A., et al.: REQUITE: a prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer. *Radiotherapy and oncology* **138**, 59–67 (2019)

25. Simkó, A., Kronsteiner, M., Glatzer, S., Vu, M., Lundman, J.A., Jonsson, J., Olofsson, J., Sandgren, K., Lechner, W., Georg, D., et al.: A physics-informed, plug-and-play dose engine for gradient-based radiotherapy treatment planning. arXiv preprint arXiv:2512.18863 (2025)
26. Thérien, B., Joseph, C.É., Knyazev, B., Oyallon, E., Rish, I., Belilovsky, E.: μ lo: Compute-efficient meta-generalization of learned optimizers. In: OPT 2024: Optimization for Machine Learning
27. Wang, S., Sun, J., Xu, Z.: Hyperadam: A learnable task-adaptive adam for network training. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5297–5304 (2019)
28. Wu, X., Yang, D., Sheng, Y., Ge, Y., Wu, Q.J., Wu, Q.: Illusion of convergence: Search space geometry in radiotherapy treatment plan optimization. *Medical Physics* **53**(1), e70209 (2026)
29. Wu, X., Yang, D., Sheng, Y., Wu, Q.R.J., Wu, Q.: A machine learning toolkit assisted approach for imrt fluence map optimization: feasibility and advantages. *Biomedical Physics & Engineering Express* **11**(3), 035016 (2025)
30. Yang, D., Wu, X., Xie, Y., Li, X., Wu, Q., Wu, Q.J., Sheng, Y.: Foresight planning: Radiotherapy plan optimization via self-supervised model predictive control. *Medical Physics* **52**(11), e70132 (2025)
31. Yin, T., Gharbi, M., Park, T., Zhang, R., Shechtman, E., Durand, F., Freeman, B.: Improved distribution matching distillation for fast image synthesis. *Advances in neural information processing systems* **37**, 47455–47487 (2024)
32. Zhang, L., Gui, Z., Zhang, P.: Aperture shape generation based on adagrad. *Nuclear Technology* pp. 1–14 (2025)
33. Zheng, B., Yang, T.: Revisiting diffusion models: From generative pre-training to one-step generation. arXiv preprint arXiv:2506.09376 (2025)